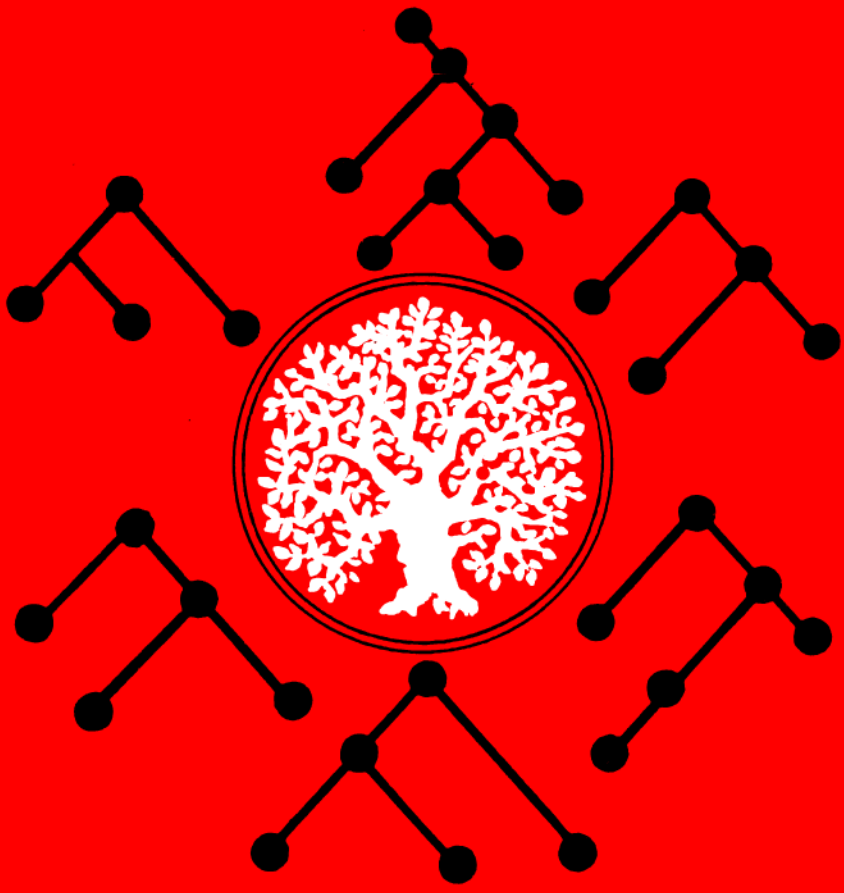
Electric
Power Systems

CYBERNETICS IN ELECTRIC POWER SYSTEMS



Mir Publishers Moscow

Electric
Power Systems

CYBERNETICS IN ELECTRIC POWER SYSTEMS

This book has been written as a study aid for students at power and electrical engineering colleges, especially where the curricula include courses on various aspects of power-system cybernetics. Main emphasis is placed on the forecasting, planning and operation of power systems. It sets forth methods and techniques for solving associated problems, including those which involve the acquisition, processing and transmission of information.

Apart from students, the book may be of value to research workers, post graduates and engineers concerned with power-system control and allied fields.

ИЗДАТЕЛЬСТВО
МИР

ЭЛЕКТРИЧЕСКИЕ СИСТЕМЫ

КИБЕРНЕТИКА ЭЛЕКТРИЧЕСКИХ СИСТЕМ

Под редакцией
лауреата Ленинской
премии
док. техн. наук,
проф. В.А. Веникова

Москва

«Высшая школа»

Electric
Power Systems

CYBERNETICS IN ELECTRIC POWER SYSTEMS

Prof. V. A. VENIKOV, D. Sc. (Tech.)
(General Editor)

Translated from the Russian
by
B. KUZNETSOV

Mir Publishers
Moscow

YU. N. ASTAKHOV, V. A. VENIKOV, YU. M. GORSKY,
D. D. KARASEV, I. M. MARKOVICH

The Greek Alphabet

$A\alpha$	Alpha	$I\iota$	Iota	$P\varrho$	Rho
$B\beta$	Beta	$K\varkappa$	Kappa	$\Sigma\sigma$	Sigma
$\Gamma\gamma$	Gamma	$\Lambda\lambda$	Lambda	$T\tau$	Tau
$\Delta\delta$	Delta	$M\mu$	Mu	$Y\upsilon$	Upsilon
$E\varepsilon$	Epsilon	$N\nu$	Nu	$\Phi\varphi$	Phi
$Z\zeta$	Zeta	$\Xi\xi$	Xi	$X\chi$	Chi
$H\eta$	Eta	Oo	Omicron	$\Psi\psi$	Psi
$\Theta\vartheta\theta$	Theta	$\Pi\pi$	Pi	$\Omega\omega$	Omega

First published 1978

Revised from the 1974 Russian edition

© «Высшая школа» 1974
© English translation, Mir Publishers, 1978

CONTENTS

Foreword 7

Introduction. Methods of Cybernetics in Large Power Systems 9

I-1. General Outline of the Power Industry 9
I-2. Power Systems — an Outline 16
I-3. Methods of the Theory of Large Systems and Management Control Systems (MCS) 23
I-4. Electric Power Systems as Complex Cybernetic Systems 26
I-5. Development of Cybernetic Methods Applicable to Operation of Electric Power Systems 28
I-6. Tasks and Courses of Action of Power-System Cybernetics 32
I-7. Optimization of Large Power System Expansion and Operation 39

Chapter 1. Expansion Forecasting 46

1.1. Economic Analysis and Tasks of Forecasting 46
1.2. Forecasting Methods 62
1.3. Modelling 79
1.4. Three Groups of Tasks in Control and Management of Power Systems 93
1.5. Planning and Forecasting in the Power Industry at Present 96

Chapter 2. Optimization of Electric Power Systems During Expansion Planning 101

2.1. Optimization of Electric Power System Expansion as a Cybernetic Problem 101
2.2. Expansion Optimization from a Mathematical Point of View 105
2.3. Simplex Algorithm 107
2.4. Simplex Method 115
2.5. Modified Simplex Method 123
2.6. The Operating-State Block of Estimation Models 131
2.7. Gradient Method 132
2.8. Dynamic Programming 135
2.9. Network Block 138
2.10. Transportation Problems 140
2.11. Branch and Bound Method 152
2.12. Cost-of-Loss Block 161
2.13. Integer Linear Programming and Mathematical Formulation of an Expansion Optimization Problem 162

Chapter 3. Criterion Analysis of the Technical and Economic Problems of the Power Industry 168

3.1. Selection Criterion for an Optimal Alternative 168
3.2. Technical-Economic Models 169

3.3. Basic Tasks of Technical-Economic Analysis 173
3.4. Use of the Theory of Similarity in Technical-Economic Analysis 176
3.5. Dimensional Units and Per-Unit Quantities 177
3.6. Economical Alternative 180
3.7. Technical and Economic Proportioning 181
3.8. Similarity Criteria for Economical Alternatives 183
3.9. Technico-Economic Stability 184
3.10. Analysis for Technico-Economic Sensitivity 185
3.11. Range of a Variable 187
3.12. Canonical Form of the Discounted Cost Equation 189
3.13. Examples of Analysis of Canonical Equations 193
3.14. Dimension Matrix 198
3.15. Parametric Similarity 201
3.16. First and Second Order Invariance 202
3.17. Discrete Constraints 204
3.18. Ellipsoids of Equal Precision 206
3.19. Functional Constraints 212
3.20. Equation with Negative Generalized Constants 214
3.21. Criterion Analysis Covering the Reliability of Service 225
3.22. Criterion Analysis of Equations Containing Exponential Functions 228
3.23. Basic Outline of Criterion Programming 230
3.24. Geometric Inequality and Minimization Problems 231
3.25. Dual Functions 234
3.26. Primal and Dual Problems 239
3.27. Fundamental Theorem of Criterion Programming 240
3.28. Procedure for Solving Constrained Problems 241
3.29. Modified Lagrangian Function 243

Chapter 4. Informational Analysis of Control and Management in Electric Power Systems 245

4.1. Information and Degree of Organization Defined 245
4.2. Basic Qualitative Aspects of Information in Control and Management Systems 248
4.3. Brief Outline of the Statistical Theory of Information 251
4.4. Hierarchy of Tasks, Goals, Algorithms and Control Centres 255
4.5. Informational Description of Large System Operation 260
4.6. Informational Processes in Electric Power Systems 263
4.7. Trends in the Development of Information Theory 276
4.8. Brief Introduction to the Information Theory of Control 280

Index 313

FOREWORD

This book has been written as a study aid for students at power and electrical engineering colleges and departments, especially where the curricula include courses on various aspects of power-system cybernetics, such as forecasting, planning, automatic power-system control, information theory, etc. It is largely based on the lectures that the authors read in 1968-1973 for students majoring in power-system operation and cybernetics at the Moscow Power Institute.

The book covers relatively new material not yet fully worked out scientifically, technologically and methodologically, particularly as regards forecasting, planning and control of large power systems. Because of its novelty and some controversial views, the subject-matter lacks uniformity of presentation, general concepts and terminology. These limitations can only be removed by further collective effort, and the authors will appreciate any suggestions from readers—students, lecturers and engineers.

The contributors to the book are V. A. Venikov (Introduction and Chapter 1), I. M. Markovich (Chapter 2), Yu. N. Astakhov and D. D. Karasev (Chapter 3), and Yu. M. Gorsky (Chapter 4). Work on all chapters was supervised by V. A. Venikov who also rendered general guidance for the team and the editing of the manuscript.

INTRODUCTION

METHODS OF CYBERNETICS IN LARGE POWER SYSTEMS

I-1. General Outline of the Power Industry

Speaking of cybernetic methods as important tools in an analysis of the paths that the power industry may take in its development and in the forecasting, planning and control of its operation, we must above all define the role that power has played in the progress of human society. Technological and, to a considerable degree, social advances of man have always been related to and, in a way, decided by the power available, the energy that man has learned to generate and which is now approaching the magnitudes associated with natural phenomena, especially geophysical and galactical processes*. At the same time, the generation of huge blocks of power** is inevitably accompanied by its losses, by discharge of ash and other pollutants into the air, disturbing effects on the runoff of rivers, and by a host of other processes all of which are acting to change the biosphere, and not always for the better.

Man is using up energy in ever-increasing amounts***. In contrast to a caveman whose daily energy consumption did not exceed 2000 kcal (less than 2.5 kWh), an inhabitant of an advanced country spends at least 200,000 kcal (or about 250 kWh) every day. The world's consumption of all forms of energy more than doubles every two decades, and in this continuous increase electric energy accounts for the greater proportion (four-fold over the same time span). As continuous is the increase in the capital (about one-third of all investments) and the labour put into power generation. Today, power generation takes up more than 60% of the manpower employed in all the producing industries. The labour force on the

---

* The capacity of the world's fixed power stations is more than 1 terrawatts (TW), which is about as much as is packed in all tides (1.5 to 1.75 TW), the thermal gradients of the ocean and Earth (about 2 TW), continuous evaporation of water from its surface (0.5 TW), and is about 1/30th of that carried in the most severe hurricane. The total capacity of all power plants, both fixed and mobile, is at least 5 or 6 TW.

** In 1972, the world's electric stations generated over 5000 TWh of energy.

*** In 1970, the world produced about 6000 million tons of equivalent fuel, which is equal to 42×10^{15} kcal or 49,000 TWh of energy. This is close to 0.6% of the total fuel reserves found to date and a little less than 0.07% of the anticipated fuel reserves on the Earth. Over the past 25 years, man has used up much more than half the total amount consumed throughout his history, or about 50,000 to 70,000 million tons of equivalent fuel.

payroll of the power industry is about 15% of the total number related in one way or another to the generation, distribution and utilization of all forms of energy. This is a very complex process closely tied in with many other processes which have a direct bearing on the status of, and progress in, all industries, transport and agriculture, and, as already noted, has an appreciable effect on the biosphere (ecological system).

Undoubtedly, the power industry has, at least indirectly, affected social progress as it is related in a definite way to the amount of energy available for use by the machines and devices that do arduous work for man*. Continuous growth in the generation of electric power and advances in its utilization have given birth to a new science and art, electrical automation, and to machines that can do some of non-creative mental work for man. Now that there is an ample supply of energy available for use in both physical and mental work**, there is everything for the methods of cybernetics to step in***. Thus, availability of energy for all forms of man's activity, including his domestic and everyday needs, is a graphic evidence of a society's development and reflects not only its economic, but also cultural level.

Huge and ever-increasing blocks of power are generated, transmitted, converted and distributed by a long chain, rather than by a single plant or installation. This chain, or more accurately system, starts at a mine (or mines) where fuel is produced, extends through carriers that take the

---

* At present, there is a wide gap among countries in cultural and economic development. This gap is related in a definite manner to the availability of energy. While some nations waste huge amounts of electricity for sheer advertising, others are lacking it even for use in the most arduous production processes, and their energy balance still heavily depends on the muscles of man and animals. Although more than 1000 GW of electric power is generated all over the world, the electric stations are distributed extremely unevenly among countries and continents. Forty-three nations with a population of 30% of the world's total use 90% of all electric energy produced; the remaining countries with 70% of the world's population have available to them as little as 10% of the total. The developing countries with about half the world's population account for not more than 7% of the world's total consumption, and their per capita figures for energy and fuel are about the same as they were at the end of the past century. This discrepancy is an illustration of the contrasts typical of capitalist society where a small proportion of a nation are in the 'have' and the huge remainder in the 'have-not' bracket. Of every three persons born on the planet, two are doomed to fall in the 'have-not' group. Indeed, the gap in per capita energy consumption between the advanced industrial and developing countries has been widening. Some 15 to 18 years ago, this gap was in the ratio 25-to-1, and now it is 36-to-1. Obviously technological progress should be directed to close this gap and to advance the progress of all mankind.

** In the USSR, this figure has increased 40-fold since 1913 (Russia's top-production year prior to the revolution).

*** As defined by Academician A. I. Berg of the Soviet Union, cybernetics in general is a science dealing with the purposeful and optimal control of complex, developing systems. Systems having or needing this form of control are called cybernetic. Cybernetics extends man's capabilities in the field of theory and analysis and in the application of their findings to practical needs in much the same way as a machine extends the capabilities of man's hands.

fuel to steam-generating units, the turbines that convert the energy of steam to mechanical energy, and ends up at generators driven by the turbines to produce electricity.

For the first time the importance of electric power and electrification as a basis for scientific and technological advances in the national economy and as social factors in general was stressed by V. I. Lenin. It was at his initiative during the early years of the Soviet state that the GOELRO plan was drawn up, the world's first integrated plan based on a scientific footing, for the economic development of the country on the basis of electrification. In precise form, this plan laid down the principles of technological policy as regards the expansion of the country's fuel and energy supplies and generation of electrical energy. The GOELRO plan mapped out a strategy based on the concentration of electric power generation and centralized power distribution through large electric power systems. At the same time, the GOELRO plan advanced the idea of unity of the country's fuel and energy supplies and of their organic ties with the national economy as a whole. Of course, in quantitative terms the GOELRO plan (which set the goal at 8.8 TWh of energy and less than 100 million tons of equivalent fuel for the country's overall energy and fuel balance) is no match to the level achieved at present. This level involves many novel scientific and technological problems of challenging complexity, and these problems are being tackled on an entirely different technical basis but in accord with the ideas first expounded in the GOELRO plan and still in force today. These ideas may be summed up as follows:

(1) Most economical utilization of fuel by pooling the generating capacities of fuel-fired and hydro stations into interconnected systems. At the time of the GOELRO plan this was called the "clustering of electric stations". To a considerable degree, this approach was more advanced than that used in the USA and Western Europe; interconnections still exist, although in a modified and improved form.

(2) Parallel operation of different types of stations to secure the most advantageous operating conditions for the system as a whole.

(3) Wide use of relatively low-grade fuels (called 'indigenous' at the time of the GOELRO plan) by electric stations for both power generation and centralized heat supply.

(4) Wide use of water power resources by construction of large hydroelectric stations and supply of electricity from water to areas lacking fuel.

(5) Construction of high-voltage power transmission lines to interconnect large stations as a way of increasing the share of electricity generated from inexpensive indigenous power-generating fuels of low heat value, which would be uneconomical to carry elsewhere.

(6) Construction of new condensation thermal power stations, preferably away from large cities, so as not to aggravate air pollution, and to cut down the costs of fuel carriage, slag disposal, etc.

(7) Improvement in the reliability and efficiency of the individual stations within an interconnected power system.

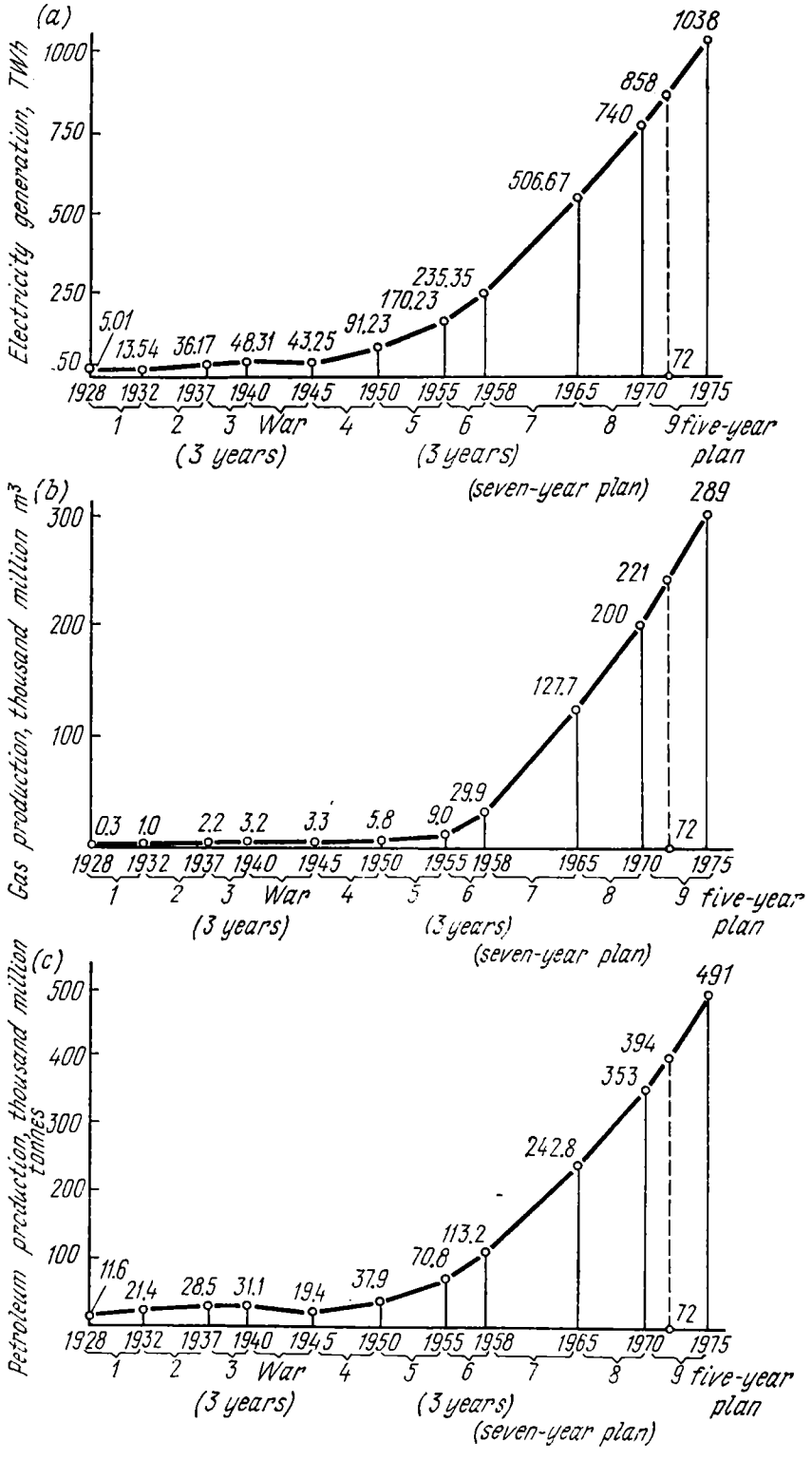

(a)
Electricity generation, TWh
1000
750
500
250
50
5.01
13.54
36.17
48.31
43.25
91.23
170.23
235.35
506.67
740
858
1038
72
1928 1932 1937 1940 1945 1950 1955 1958 1965 1970 1975
1 2 3 War 4 5 6 7 8 9 five-year plan
(3 years)
(3 years)
(seven-year plan)
(b)
Gas production, thousand million m³
300
200
100
0.3
1.0
2.2
3.2
3.3
5.8
9.0
29.9
127.7
200
221
289
72
1928 1932 1937 1940 1945 1950 1955 1958 1965 1970 1975
1 2 3 War 4 5 6 7 8 9 five-year plan
(3 years)
(3 years)
(seven-year plan)
(c)
Petroleum production, thousand million tonnes
500
400
300
200
100
11.6
21.4
28.5
31.1
19.4
37.9
70.8
113.2
242.8
353
394
491
72
1928 1932 1937 1940 1945 1950 1955 1958 1965 1970 1975
1 2 3 War 4 5 6 7 8 9 five-year plan
(3 years)
(3 years)
(seven-year plan)

Fig. I-1

The further development of the country's power industry under the decisions of the 24th Congress of the CPSU ensured the generation of 1038 TWh in 1975 (the closing year of the 9th five-year period) and the increase in the capacity of power stations to 218 GW (of which amount, 150 GW is to be supplied by the country's integrated supergrid). This increase in generating capacity will mainly be due to thermal power stations using large generating units. It is to be noted that between 1971 and 1975 the country's generating capacity increased more than 1.4 times in comparison with the 8th five-year period. The rate of growth is illustrated in Fig. I-1.

Under the 9th five-year plan a good measure of expansion was planned for nuclear power, including nuclear power plants with reactors rated at 1000 MW and more each and with a total power output of 6000 to 8000 MW.

Appreciable headway has been made in the construction of the USSR Integrated Power Supergrid*. Towards the end of the 9th five-year period, the total length of overhead power transmission lines operating at 35 kV and higher was to rise to 603,000 km against 446,000 km in 1970 and 124,000 km in 1960. The breakdown of electricity consumption by the various divisions of the national economy is shown in Fig. I-2. Quite naturally, the biggest consumer is industry, although a good share of the total is set aside for agriculture. In 1970, the country's 99.8% collective and state farms used electricity and drew a total of 38.5 TWh from the state-run power systems. Under the 9th five-year plan, nearly twice as much energy was to be made available to agriculture (including the needs of the rural population). A sizeable increase in electricity consumption was also planned for railways. By the close of the 9th five-year period, some 40,000 km of railways were to switch to electric traction, which would amount to 28% of the total length, and to use 75 TWh overall (see Fig. I-2).

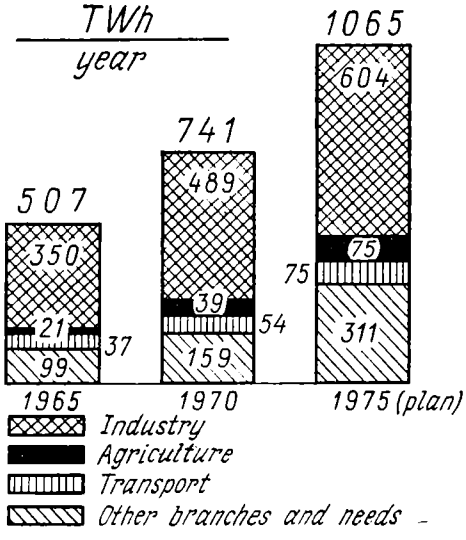


Fig. I-2

On the qualitative side, advances were planned to be made not only in the construction of new power generating capacities, but also through a systematic improvement in the performance of power-generating plant. For example, in 1975 the fuel consumption per kWh at the electric stations was to be cut down to 340-342 g. Assuming a decrease of 24 to 26 g per

* The plan figures quoted here and elsewhere have mainly been achieved. Towards the end of the 10th five-year period (1980) it is planned to generate 1400 TWh and to build power stations with a capacity of 70 GW.

kWh delivered, the economy for the whole of the country over the 9th five-year period was to be 50 million tons.

Labour productivity in the electric power industry was planned to increase 1.4 times in 1970-1975. If labour productivity had been kept the same, the production of electric and thermal energy planned for 1975 would have called for an additional labour force of 232,000 under the USSR Ministry of the Power Industry alone. With the planned increase in labour productivity, mainly through technological progress, it was expected that the added labour force would not exceed 25,000.

This progress of the power industry was possible owing to the achievements made in the preceding, 8th, five-year period. In 1966-1972, the production of electric energy rose from 507 to 858 TWh a year, and the capacity of all electric stations in the Soviet Union as of January 1, 1972 was 1783 GW.

Under the decisions of the 24th Party Congress, work was continued on the Integrated Power Supergrid for the European USSR, with a total capacity of 110 GW. The Supergrid was to incorporate 53 power systems operating in the European USSR and the Urals. This is the world's largest power grid run from a single control centre. Under the 9th five-year plan, its capacity was to be increased to 153 GW (by 46%). The Siberian Integrated Power Grid, with a capacity of 22.8 GW in early 1971, was to deliver 29.1 GW (an increase of 28%) in 1975.

At the time of this writing, the power industry had over a thousand dispatching and control centres operating on a round-the-clock basis at various levels of the centralized hierarchical system. The USSR Ministry of the Power Industry operated a ramified network of communication channels for the transmission of control and, to a lesser degree, economic data, including 800,000 channels using power transmission lines, 200 radio-relay stations, and over 20,000 VHF/UHF and HF radio stations.

With a relatively small number of digital computers in use at various levels of the dispatch hierarchy, and continually updated information, it is now possible to distribute load in the most economical way among all the main power stations of the Integrated Power Supergrid.

Work is under way to set up an integrated management information system (IMIS) based on interconnected computer control centres, of which there will be one master and eleven slaves. Some of them were to be put in operation under the 9th five-year plan.

A distinction of power science in the Soviet Union is the idea that the country's power (fuel and energy) system is an indivisible entity encompassing everything between energy resources and utilizing plant, an idea which was first laid down in the GOELRO plan. It was at that time that the so-called pan-energy method of research was formulated to be later widely used in both the theory and practice of power science and engineering. In the late 1950s and early 1960s, this method, taking a new dimension owing to cybernetics, was formulated as the systems approach to problems of power engineering.

Power science made a new qualitative jump with the advent of cybernetics and the theory of large systems*, and especially with the development of computers and computational mathematics. By the early 1960s, fundamental and applied research in the optimization and control of large power systems had become an urgent necessity in the light of the problems raised by the continuous progress of the power industry. It was at that time that early steps were made to set up the country's pan-energy system, a unique large system (embracing all the basic elements and connections of the country's energy balance), on the basis of ever-closer connections between, and ever-increasing interchangeability of, the various forms of energy and energy resources.

In recent decades, interconnected electric power systems have been growing into nation-wide integrated pan-energy systems incorporating, apart from electric power systems, also integrated heat, gas and oil supply systems whose connections with all divisions of the national economy are growing ever tighter. These interrelated large systems are becoming still more ramified and complicated, as may be seen from the chart of Fig. I-3**. As they grow and expand, the individual large subsystems are inevitably merging into an organically interwoven large management system. Ever more stringent requirements have to be satisfied as regards the reliability of all elements of this system, and especially its electric power and gas supply sectors, and one is forced to treat as a single package the problems involved in the choice of an optimum system structure at the design stage and during system operation and control.

Thus, Lenin's concepts of electrification have been finding embodiment in both the establishment and development of power systems. His concepts underlie not only improvements in the production and distribution of electric energy, but also the continual expansion of electrification, the optimum level for which is going up all the time, in step with advances in science and technology. As a consequence, whatever is done to develop electric power systems, it is vitally important to foresee, on a scientific basis, the future needs of the national economy in electric energy, that is, to forecast this development on the basis of an objective evaluation of the levels of electrification achieved and those taken as optimal.

---

* For the purpose of our discussion, a large system will be defined as that which is made up of a number of subsystems with a large amount of feedback such that they cannot be isolated from one another, and whose behaviour is decided by the interaction of these subsystems. Thus, the power industry is a large man-made cybernetic system (for a definition of cybernetics, see p. 10).

** The hierarchy (at the top of Fig. I-3) has been proposed by Academician L. A. Melentyev. The diagram shows both vertical and horizontal connections. The abbreviations stand as follows: IEPS, integrated electric power system; IGSS, integrated gas supply system; IOSS, integrated oil supply system; ICSS, integrated coal supply system. The solid lines (————) mark pan-energy connections; the dashed lines (— — —) designate connections for electric power; the dash-dot (— · — ·) lines show connections for oil and oil products; the dash-double dot (— · · —) lines, connections for coal; and the dotted lines (. . . .), for natural gas.

Today, different forms of energy and energy resources can be used to satisfy the needs of the national economy. Therefore, an estimate of optimal levels for electrification can markedly affect the production of natural gas, oil and oil products, coal and heat. This is an extremely complex problem of vital scientific and practical significance, and the way in which it is solved will fix the road for the development of the basic types of large power systems. In turn, if the solution is to be sound, it is important to estimate correctly what the national economy and the consumer stand to win from electrification.

In the early years of electrification, effort was mainly put into stationary power and lighting plants both in industry and in the homes. Later (starting with around 1940s), emphasis was placed on the use of electricity in high-temperature manufacturing and chemical processes, along with its uses for domestic needs and on railways. The closing decades of the 20th century will probably see an increasing use of electricity for heating, air conditioning and propulsion (notably in electromobiles). This is an indication of a steady rise in the use of electricity in the country's pan-energy balance (about 3% in the 1920s, 15% at the time of writing, and close on 40% towards the year 2000), and of the fact that the power industry has been expanding at a rate faster than the average rate of the national economy as a whole. The prospects are that in the nearest forty to fifty years, electricity will become the basic form of energy for use in the national economy. This is why the optimal development and control of the country's electric power systems has become a major problem both scientifically and practically, and cybernetics, the science of control, assumes a special importance in this field, too. It is the objective of this book to bring out some of the problems of cybernetics as applied to power generation and utilization.

I-2. Power Systems — an Outline

For the purpose of our discussion, we shall apply the term electric power system to the electrical part of an interconnection of electric power stations linked by power transmission and distribution lines and load centres into an *integrated productive complex*. Several individual electric power systems may be, and usually are, combined into an interconnected electric power system. An interconnected electric power system, when it covers all or a sizeable portion of the country's territory, will be called an *integrated power (super-) grid* (see Fig. I-3)*.

Electric power systems may differ in *structure*, that is, in the mix of the constituent stations (which may be condensation thermal power station, power and heat generating stations, hydro-power stations, nuclear stations, etc.).

* Electric Systems, vol. 1. Edit. by Venikov V. A. "Higher School" Publishers, 1970; Markovich I. M. Operating States of Power Systems. "Energiya", 1969.

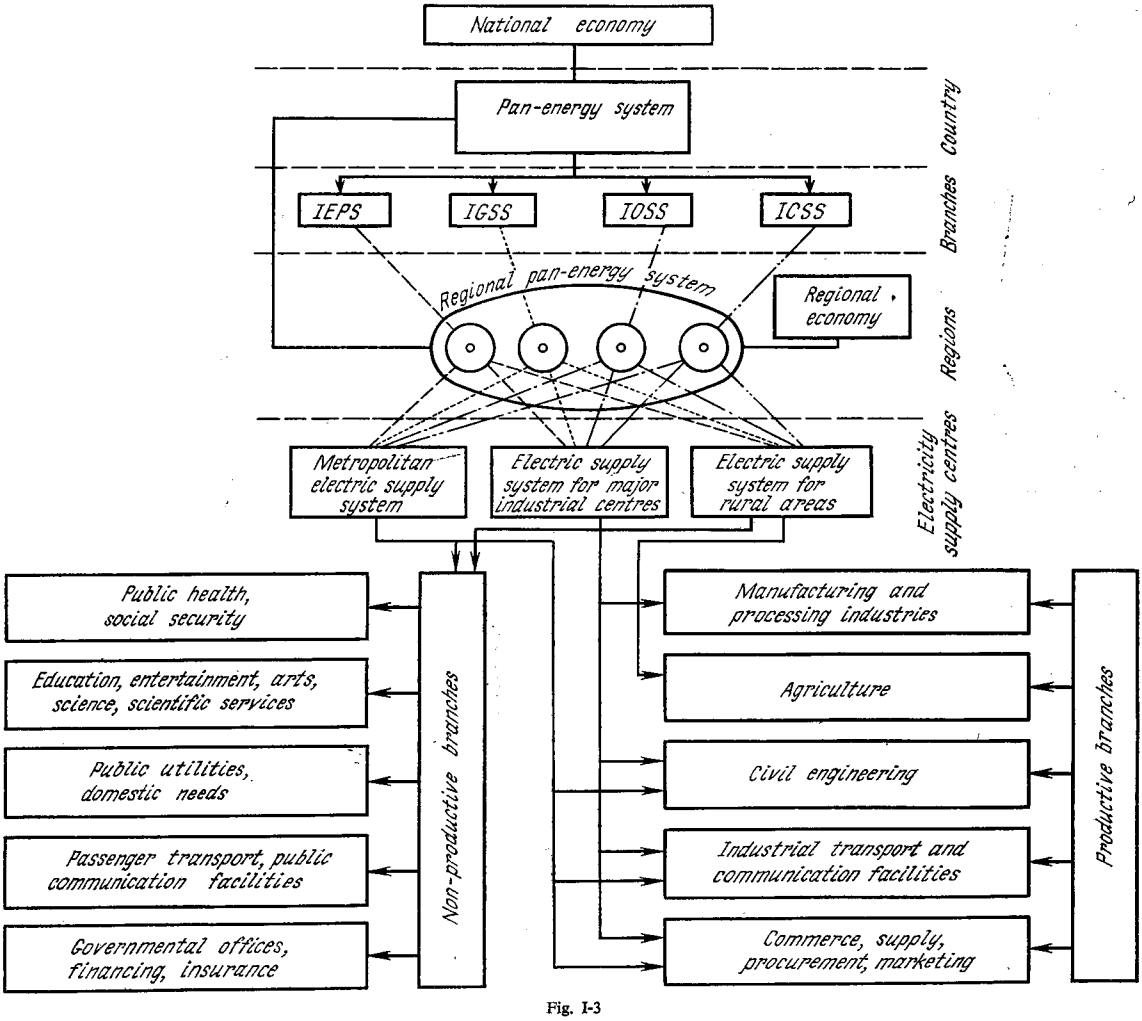

National economy
Pan-energy system
IEPS
IGSS
IOSS
ICSS
Regional pan-energy system
Regional economy
Metropolitan electric supply system
Electric supply system for major industrial centres
Electric supply system for rural areas
Country
Branches
Regions
Electricity supply centres
Public health, social security
Education, entertainment, arts, science, scientific services
Public utilities, domestic needs
Passenger transport, public communication facilities
Governmental offices, financing, insurance
Non-productive branches
Manufacturing and processing industries
Agriculture
Civil engineering
Industrial transport and communication facilities
Commerce, supply, procurement, marketing
Productive branches

Fig. I-3

Electric power systems may also differ in *configuration* and *connection scheme*. The term configuration will apply to the manner in which the constituent stations and interconnecting networks are diagrammatically arranged within a given system, and the term connection scheme will define the tie-ins, also in diagrammatic form, between the constituent stations and the associated load centres.

Individual electric power systems may be linked together by *inter-system connections* to facilitate interchange of power between them.

The transmission lines, or interconnectors, terminate at substations. There may be substations to step up or down the voltage transmitted (step-up or step-down substations).

Electric power is conveyed to consumers over *electric networks*. An electric network is made up of lines, substations, distribution and switching centres. In the Soviet Union, all electric networks are classed into *cross-country networks* which carry electricity from generating stations to main substations, *regional networks* each supplying power to a limited area and *distribution networks* which are parts of a regional network and may have service connections to consumers at any point. Networks may also be classed into *feeders (*or *bulk-supply networks)* which are not connected directly to utilizing equipment, and *distributors* which supply loads in urban, rural or industrial areas.

At any particular instant, an electric power system may be at a particular *state* which may be defined by a set of events or processes that are typical of the system's state. In turn, these events or processes may be described in terms of *state variables*, such as power, voltage, current, frequency, etc.

As a physical entity, an electric power system may be described in terms of *system variables* which are dependent on its structure, configuration and connection scheme, and which in turn govern the coefficients or factors that relate the state variables.

A power system may be in a *steady* or a *transient state*, or *condition*. Under steady-state conditions, the state and system variables are practically constant, but they vary during transients. As a further distinction, a power system may be in a *normal state of operation*, supplying the normal quality of power and developing no faults. In contrast, it may be in an *abnormal* or *fault state*, or *condition*, arising either from a fault or revealed by a poor quality of energy delivered. Finally, a system may be in a *restorative* or *post-fault state*, that is, a condition immediately following the clearance of a fault and restoration of energy supply or services.

As already noted, a pan-energy system is made up of subsystems, such as electric, fuel, distribution, gas supply (which may in turn be interconnected), coal supply, oil supply, nuclear fuel supply, etc. (see Fig. I.3).

Typical pan-energy systems are global — they cover large portions of a country's territory or even several countries. This is why they are large in size and show extremely complex relationships which are specific in quality because of the quantitative aspects involved.

It is likewise typical of pan-energy systems that both the production and utilization of energy is continuous and inseparable, because firstly their products are widely interchangeable and, secondly, what is output from one system serves as input to another.

Pan-energy systems bring their influence to bear on both the national economy and technological progress in at least two important directions. For one thing, their growth depends on demand for energy and on the materials and labour that a nation can set aside for power industry development. For another, the power industry and, especially, electrification can directly shape technological progress, the development and distribution of productive forces.

All of these factors, falling in the category of external connections between energy systems and the national economy, can be investigated within the framework of a theory of, and incorporated into methods for, the control of pan-energy systems as large systems. Pan-energy systems may be approached from a *physical and engineering angle*, that is, from the viewpoint of their basic material relationships, the electrical, gas and oil supply networks involved. Alternatively, they may be viewed from an *economic standpoint*, with the material relationships largely reduced to purely informational form. The physical and engineering approach comes in where one is to handle operational control tasks, and the economic approach, in the control of system expansion. Of course, both aspects are closely interrelated, and this, too, is a specific feature of energy systems.

The crucial problem in the control of energy-system development and operation is the need for appropriate methods. In this connection, use may be made of the following means of control and management: (1) models to investigate and improve control and management functions; the models may be physical, analog, and mathematical, realized mainly on digital computers; (2) electronic control hardware to run the processes involved, to gather, process and transmit data; and (3) organizational, that is, economic, managerial and incentive means and methods the ultimate goal of which is to guide the personnel participating in system operation and control.

The rational control of a pan-energy system as an assemblage of large systems reduces basically to finding *optimum combinations* of the three groups of means listed above, in various hierarchical (technological and territorial) and time conditions. It is also to be remembered that man-made large systems actively interact with the environment.

Thus, the operation (expansion) of a large system proceeds continuously through a combination of actions by man and machine, and feedback. A control system must be fully aware of this aspect of the *controlled* energy system.

In a pan-energy system the basic connections between fuel and energy supplies are those between the subsystems of an integrated gas-supply system, an integrated oil-supply system, an integrated coal-supply system, and an integrated nuclear-fuel supply system. There is, however, a different

classification under which each of the integrated electric power, gas-supply, oil-supply and coal-supply systems is treated as an independent large system closely related to the other systems and divided on a geographical basis, each subdivision having an integral hierarchy of partly independent large systems*.

System control and management should be examined at three levels as follows: (1) operational control (dispatching) over days and seasons of the year. It is affected by the duration of transients (which may last for a split second in an electric power system and for up to several minutes or even hours in a heat or gas supply system); (2) management on a yearly basis; and (3) expansion management which may cater for periods of 5 years, 10 to 15 years, and 20 to 25 years ahead. At each of these levels, there are three factors that define a system as the object of control, namely technological, geographical, and temporal.

Integrated pan-energy systems are closely related to the national economy by external connections. These connections are (1) external energy connections which characterize the utilization of the various forms of energy and energy carriers by the national economy and the accompanying 'non-energy' payoff; (2) external material, labour and economic connections or flows which characterize the needs of the fuel and energy producing industries in equipment, raw materials, manpower, capital, and running expenses. Practically, the external connections, or flows, of the fuel and energy producing industries with the national economy can be defined between limits chosen rather arbitrarily. As a rule, the values assigned to these connections or flows are appreciably in error, and the corrections applied during the subsequent optimization of system development may result in anything but optimum conditions for the national economy as a whole (for example, the national economy may suffer a loss if, in optimizing the operation of hydro-power stations, only the gain from power generation is considered, however accurate it may be, but the other users of water resources are neglected).

As large systems, pan-energy and electric-power systems show a number of specific qualities:

(1) their development in time is subject to constraints mainly decided by external connections;

(2) they show a multiplicity of nonlinear relations; this stresses the fact that the basic relations within each system are fundamentally nonlinear and of a mass scale;

(3) any changes in state variables and actions in a system are probabilistic in the sense that the system state at some succeeding instant cannot be uniquely deduced from its previous state, mainly because the human factor in the system defies accurate evaluation and also because there may be quite a number of random disturbances;

* For a detailed discussion, refer to Melentyev's L. A. "On the theory of control of large systems". USSR Academy of Sciences Press. "Energetika i transport", 1969, No. 4.

(4) the systems are self-regulating in the sense that they can adapt themselves to changing conditions, both external and internal;

(5) every system incorporates feedback which is a major contributor to the system's ability for self-regulation;

(6) the systems are extremely reliable owing to redundancy of elements and connections.

It is important to isolate the properties, or behaviour, of real man-made large systems as controlled systems, from the tasks and goals sought in controlling them.

The ultimate goal of controlling a man-made large system is to maximize some performance, or criterion, function subject to certain constraints*. To achieve it, two different, but interrelated classes of tasks need be tackled:

(1) *Operational tasks*, basically reducing to:

(a) the selection of the most efficient system structure, and from it, the most appropriate means, methods and organizational forms for its control;

(b) planning the system operation (expansion) for different spans of time (which are likewise to be chosen on an optimum-seeking basis);

(c) an analysis of the conditions in which the system is to operate (grow) within the time intervals chosen;

(d) review and adjustment of the plans for system development in the light of the above analysis;

(2) *Implementation tasks* which are basically the implementation of the plans and strategies chosen at the previous stage. This involves the collection, transmission and processing of the necessary information, analysis of likely disturbances and generation of corrective or restoring actions.

The ways and means involved in handling operational and implementation tasks and the tasks themselves are related, on the one hand, and substantially different on the other in different types of man-made large systems where the various parts may be coupled by electric, electromagnetic, mechanical or informational connections. For example, all elements of an electric network operating at the same voltage level are coupled electrically; networks operating at different voltage levels are coupled electromagnetically by transformers; the stators and rotors of generators and motors are similarly coupled electromagnetically. Prime movers are coupled to generators and motors to driven machines by mechanical means; boilers and steam turbines by pipelines, and water reservoirs to hydraulic turbines by penstocks. All of these connections transfer energy in one form or another.

The development of the power industry in the form of pan-energy systems is not only more economical, but also more reliable as this improves the continuity and security of service and utilization of resources (such as the waste gases of blast furnaces, wind and solar power) which would otherwise be wasted or left untapped. Pan-energy systems offer a convenient

* For a detailed discussion, see *Engineering Cybernetics in the USSR*, "Nauka" Publishers, 1968.

means for the joint generation of different forms of energy — electric, thermal, etc.

It is relevant to sum up* some of the features of electric power systems most important to their control:

(1) Since its generation, distribution and utilization take place at the same time, no energy can be generated if there are no consumers to use it — generation is uniquely decided by demand, and vice versa. It is to be noted that energy is converted and transmitted with some losses in all elements of a system, and this factor must be taken into account. In view of the foregoing, the following corollaries may be stressed as regards electric power systems:

(a) any reduction in generation at electric stations (because of scheduled or unscheduled equipment outage, faults, or other causes) immediately leads to a reduction in the quantity of energy delivered to consumers, if there are no sources capable of making up for the deficiency (reserve capacities);

(b) a temporary reduction in demand because of the consumer's plant outage, faults or other causes results in the waste of the power-generating capacities at electric stations, if the consumer's system has no regulators to cope with the contingency;

(c) no imbalance can exist between the total generation by electric stations and the total demand by the system. As generation goes down, the power utilized does the same, and vice versa; this may be accompanied by an impairment in the quality of power delivered.

(2) All transients in an electric power system are fast, and this necessitates the use of suitable automatic facilities. These devices, often with a very high speed of response, must provide for normal course of transients in the system.

(3) The fact that electric power systems are inseparable from the other divisions of the national economy stresses the need for timely system expansion and growth. In fact, the electric power systems must grow faster than power demand since otherwise the systems would be unable to keep any capacities in reserve.

Electric power systems may be classed in several ways. On the basis of the energy resources used and, as a consequence, the types of electric stations included, existing electric power systems may be classed into:

(a) systems with thermal power stations which utilize the energy of fuel, including nuclear fuel;

(b) systems with hydro stations which utilize water power;

(c) mixed stations with both thermal and hydro power stations.

On the basis of the consumer mix, there are:

(a) systems which serve lighting and domestic needs;

(b) systems which energize industrial motors;

(c) systems which supply tractive power;

* After Markovich I. M. *Operation of Power Systems.* Energiya, 1969.

(d) systems which energize various furnaces, kilns, and similar plant;
(e) mixed systems.

Still another way of classifying electric power systems is on the basis of the relative geographical position of power generating stations and load centres:

(a) concentrated, where the stations are located close to load centres and there is no need for long-haul power transmission;

(b) extended, where the energy resources and the stations utilizing them are distant from load centres and the need therefore exists for long-haul transmission lines and ramified networks.

Electric power systems may be interconnected by means of single- or multi-circuit transmission lines. Basically, interconnected power systems offer the following advantages:

(a) a reduction in the amount of plant that need be held in reserve;

(b) better utilization of the power and energy generated by the hydro stations of one or several systems, and improvement in their economy;

(c) a reduction in the total peak load to be supplied by the interconnected systems;

(d) assistance between systems in the case of seasonal variations in station capacity especially in that of hydro electric stations;

(e) assistance in supplying seasonal variations occurring in one system by energy transferred from another where they occur at another time;

(f) assistance between systems at the time of scheduled or unscheduled outages.

As electric power system interconnections continue to grow and expand, it becomes more and more economical to merge them into an integrated power supergrid supplying power to large parts of the country*. In addition to the advantages arising from interconnections, those offered by an integrated power supergrid may be stated as follows:

(a) the 'longitude' effect taking place where the systems or interconnections pooled into the integrated supergrid are at widely differing geographical longitudes, and the peak loads occur at different times of day for different constituents of the grid (load diversity). Because of this, a smaller amount of spinning reserve capacity is needed at each station, and their construction cost can therefore be cut down appreciably. Power interchange over the system interconnectors and tie-lines may vary considerably or even be reversed at peak-load hours, because the constituent systems utilize the tie-lines to assist one another in supplying peak loads;

(b) the 'latitude' effect which takes place where the constituent systems are at widely differing geographic latitudes so that the peak load duration is different and systems within the supergrid with a shorter peak load duration can assist others by supplying their peak loads;

(c) intervening power systems can readily be connected to the supergrid, and the intervening areas can advance their electrification at a lower cost.

* See the footnote on page 21.

The most important problems in connection with the integrated power supergrid are, firstly, the use of completely automatic control to maintain operation of the constituent stations, networks, systems and interconnections under normal conditions and continuity of service under abnormal conditions; secondly, control of power interchange over the supergrid tie-lines so as to ensure the overall economy and reliability of the supergrid and of its elements.

It is one of the ultimate goals of planning to choose the most economical path for the development of a power system in terms of money invested and services obtained.

Currently, a development plan for a power system will be regarded as an optimum one under a specified set of conditions, if it minimizes the discounted cost criterion. However, the minimum discounted cost can no longer serve as the only criterion with regard to similar plans for the development of interconnections and the supergrid, because the same individual power systems and tie-lines are now treated as an integral whole. An increasing emphasis is being placed on reliability and environment protection, and the objectives of optimization may therefore be entirely different.

Mathematically, an optimization problem reduces to minimizing a very complex nonlinear function of a great number of variables under constraints in the form of equalities and inequalities, with some of the variables being integers. This is a dynamic multiextremal problem which cannot yet be solved by existing techniques.

I-3. Methods of the Theory of Large Systems and Management Control Systems (MCS)

As regards their development and use, the methods of power-engineering cybernetics are closely related to the theory of large systems and to the methods and means of automatic planning and management, notably to the establishment of the State Computer Centre Network. Therefore, it will be worth while to discuss these general matters of economic management as a whole before we take up the more specific problems of power-system cybernetics.

The management of the national economy makes it imperative to advance the theory of production and economic systems, and to perfect and more widely use systems analysis. Fundamental research in the field of large systems and systems analysis by cybernetic methods with reference to the national economy in general and to the power industry in particular must be conducted jointly with mathematical economic analysis, optimum planning, and social and psychological studies in the field of management. In view of these requirements, four cardinal directions can be singled out in the theory of large systems.

Work on principles of economic management and the tasks of power-system cybernetics. If the national economy is to grow and operate normally, it is vitally important to formulate precisely the goals and tasks for

the appropriate control and management systems. A realistic formulation is especially important with regard to major projects, such as space exploration, exploitation of off-shore areas, development of the Arctic region and Siberia. This class of projects also includes the development of energy and raw-material supplies. Further improvements in the planning and management of the national economy, including the establishment of the State Computer Centre Network, will undoubtedly help with these projects. What is especially important as regards guidance and management in this field is to concentrate scientific effort on the vital and decisive directions. In the formulation of goals and tasks it is important to take into account the specific features of man-machine systems making up a multi-level hierarchy, each adapted to handle particular tasks in particular fields. The search for optimum courses of action and evaluation of final results will need appropriate criteria based on this hierarchy of goals and realization time. The hierarchy (tree) of goals (or tasks) weighted in terms of importance and mutual utility must provide a basis for research, on the one hand, and serve as a tool with which advances in science and technology could correctly be predicted, on the other.

Programs for the management of the national economy and of its divisions as complex, expanding systems of a cybernetic type. Economic management on the branch (sectoral) basis is advantageous in that it enables each industry, or sector, to pursue a policy best suited to its needs, improve labour productivity through specialization, and utilize scientific advances specific to that industry so as to promote technological progress and computerized management.

The negative aspects of the sectoral form of economic management, such as the use of local optimality criteria unrelated to any global criterion (parochialism), the tendency to an isolated and self-sufficient development (parallelism), can be neutralized through the use of the so-called program management method.

With this method, the activities of the individual industries, or sectors, can conveniently be correlated towards the achievement of goals or tasks common to the national economy or the nation as a whole. These nation-wide goals and tasks may be those of a global magnitude, formulated in programs or five-year plans, or they may be a multitude of specialized projects, such as an accelerated expansion of road transport on an integrated basis, control of water and air pollution, research projects in the field of geophysics, etc. In any of these instances, several industries or sectors may have to pool their efforts in a coordinated manner, and the solution cannot be limited to any particular branch of the national economy. In the conditions of limited resources, it is important to allocate capital to a particular industry in proportion to what it can contribute to the attainment of the common goal. However, as the national economy continues to grow and expand, it becomes increasingly more difficult to coordinate the activities of the individual industries and to allocate capital investments. On the one hand, progress in each particular industry must be coordinated

with that in any other so as to lead to the common goal in the most efficient manner; on the other, discrepancies may well occur between what the individual industries contribute to a nation-wide goal and what has been invested in them to promote their progress, thereby interfering with the optimal utilization of available resources. As a result, disproportions may develop in the national economy and the targets may be achieved with delays. This is where development programs, each covering a logically interrelated set of activities in production, management and science in the various industries, can help to eliminate such discrepancies and delays. For instance, a development program for the automobile industry should provide for the expansion of not only automobile production, but also for that of motor-roads, motels, car-service centres, and the like. Also, such a program should have provisions to train personnel and to carry on research and development. As an example, the development program for Siberia covers in a coordinated manner the expansion of industry, agriculture, timber, power generation, roads, communications, city planning, tourism, labour force, scientific facilities, etc. A similar approach is taken to drawing up other programs — in all of them, a nation-wide goal is predominant and each takes into account the feedback from the other industries. That is, matters are handled by methods and techniques of cybernetics.

Automation in the management of the national economy. Methods of automatic control and, in the long run, those of cybernetics must develop in step with the methods used to analyse, synthesize and model the structure of economic subsystems and along with the theory of systems and automata in the development of formal algorithmic languages. Also falling in the category of analysis and synthesis are matters related to planning and decision-making.

As regards economic management, it would be important to investigate organizational structures of the hierarchical type. These investigations should cover both the *branch* (sectoral) and the *goal* (task) *hierarchies** associated with particular programs and subprograms. The point is that the pressure from different sectors and different goals (tasks) poses a variety of delicate problems before planners and decision-makers concerned with the development of large systems and complexes involving the participation of hundreds or even thousands of organizations under the jurisdiction of different industries and authorities.

Optimization and theory of optimum decision-making. Among the mathematical tools of decision-making are the theory of games, waiting-line (queueing) theory, combinatorial mathematics and mathematical programming. These methods are decisive in analysis and control of large systems in general, and especially where this is done with the use of cybernetics and

* Makarov A. S. et al. *On the Spatial and Temporal Hierarchy in the Long-Term Planning of Electric Power Systems.* Collected volume "Mathematical model", Irkutsk, 1971.

operations research. It is important to develop multi-step optimization and appropriate multi-step decision-making procedures leading to the optimum functioning of the national economy and providing a sound basis for the use of digital computers and computer systems. Special emphasis must be placed on statistical methods, notably those giving estimates of reliability and efficiency already at the design and development stages. In the field of statistical theory, it would be important to use various combinations of statistical techniques based on mathematical and physical models (including physical simulation, or partial system and subsystem tests), with special reference to the running, maintaining and restoring of large systems. The 'man-machine' relations, too, should receive proper attention within the framework of large systems.

In both large-system analysis and synthesis, emphasis should be gradually shifted from integrated automation to purposeful control under conditions of changing objectives, that is, to cybernetic methods.

I-4. Electric Power Systems as Complex Cybernetic Systems

It has been shown that an electric power system is related to several subsystems which are in part its constituents and in part interact with it. It has also been mentioned that these subsystems have feedback paths, which fact reflects the influence that one subsystem has on its neighbours. In the circumstances, purposeful control imparts a large electric power system the qualities of a cybernetic system. Likewise, in the chain of elements taking part in the generation, transmission and distribution of power, we can isolate individual subsystems, such as a fuel-supply system, a transport system, and a power system proper which may in turn be broken down into a thermal-power and an electric-power utilization system. Thus a power system should be treated as a complex system made up of subsystems with feed-forward and feedback. Furthermore, the constituent subsystems interact with many other large systems, such as the biosphere (ecological system), industry, transport, agriculture, etc. Also, power subsystems are affected by demographic and other factors related to society as a whole. As the power system grows and expands, the association between the subsystems become increasingly closer, and anything but optimum control in any (say, power) subsystem would immediately tell on other subsystems. At the same time, this growth and expansion may lead to interrelations between the power system and other systems, which fact, in default of optimum control, may have undesirable consequences already today. In fact, such a situation might well be catastrophic in the remote future in view of the effect that power generation and utilization has on the biosphere. Among other things, the uncontrolled expansion of power systems all over the world (using all sources of energy) to a level where they can generate 3 to 5% of the energy that the Earth receives from the Sun might upset the heat balance of the planet, causing the glaciers to thaw and the ocean to rise so much that all or nearly all land would be flooded. Such

huge amounts of energy*, comparable in magnitude with that coming from the Sun, could, for example, be produced from the controlled thermonuclear reaction (fusion of light-weight nuclei). Although this appears to be feasible in a fairly remote future and at present man has neither means nor needs for the generation of so much energy, it is quite obvious that already now the power industry must be viewed as a large, complex cybernetic system closely related to the country's economy and very existence.

Cybernetic methods can be applied to the solution of a huge number of problems such as the siting of power stations throughout the country, an optimum correlation between the capacities of power systems and their elements, on the one hand, and the land allocated to them, on the other**, optimization of station performance (not only in terms of power output and cost, but also with regard to the utilization of water*** by other users), control of air pollution**** which might take on catastrophic proportions unless appropriate measures are carried out*****.

---

* On the average, the Sun supplies 1.35 kW/m^2 of power. If we multiply this figure by half the projection of the Earth's surface area, $\pi(6.4\times10^6)^2=1.29\times10^{14}$ m^2, and by the number of hours in a year, 8760 hours, we shall find that the annual input of energy from the Sun is 0.75×10^{18} kWh, which is about 150,000 times the annual generation of energy all over the world (5×10^{12} kWh, see the footnote on page 10). The energy available from all forms of fuel produced on the Earth today (see the footnote on page 9) is less than 0.01% of the power from the Sun.

** The increase in unit capacity of stations and of their generating plant is a necessity. If the stations retain their existing unit (average) capacity, the capacities to be put in operation between 1975 and 2000 would occupy an area the size of Belgium.

*** The thermal power stations (those burning fossil and nuclear fuel) expected to be in operation in the USSR by the year 2000 will need some 40,000 m^3 of water every second, which is about one-third of the annual average run-off of all rivers in the Soviet Union. The reasonable way out appears to use closed circuit cooling systems, to build bigger cooling towers, etc.

**** The power industry (mainly, fuel-fired stations) at present is a leading, though not the worst source of air pollution. According to US data, the power industry accounts for 25% of all pollution, the remaining 75% being split between automobiles (50%), industry (20%), waste incineration and other sources (5%). Unless appropriate measures are taken, the power industry's contribution to air pollution might increase due to discharge of ash and other wastes. Today, thermal electric and other power stations are sited so that power output density and population density occur in about the same areas, with the most densely populated regions receiving the largest share of pollution. This is why the siting of thermal power stations and their capacity should be chosen by considering the power industry as part of a large system, that is, on the basis of the cybernetic approach, rather than from the interests of the power industry alone.

***** According to US sources, air pollution is to increase by another 30% towards the year 2000, and this will markedly affect the climate. Progressively increasing quantities of effluents, including the condensate from power stations, are discharged into lakes and rivers and raise their temperature. This (especially if the condensate is contaminated) may kill plants and animals. The rise in the temperature of water reservoirs, along with the large quantities of heat lost to the air, usually impairs the microclimate in the areas around the power stations. The dams across the rivers slow down the flow of water, collect dirt, and allow blue-green algae to suppress all other life: with floods no longer occurring as they would normally do, meadows go dry and the soil becomes salinated. Thus, again, we can see that the power industry cannot be looked upon as an independent entity and assessed in terms of only the energy it can supply.

A matter of equal importance is the reliability of power supply in the urban and rural areas and throughout the country. Here, too, sufficient reliability can be obtained only if the power system is treated as part of a larger cybernetic system which incorporates as its constituents the country's industry, transport, rural and metropolitan consumers. Apart from purely economic considerations, the reliability of power supply must also be viewed from a psychological angle. A major power failure in November, 1965 left New York and an area of 200,000 km^2 around it without electricity for 14 hours. This entailed losses which can hardly be measured in terms of money, and the inhabitants of the cities and towns lying in darkness were completely demoralized.

The foregoing stresses still more the importance of treating an electric power system as a large, complex cybernetic system. This approach is valid as regards forecasting, the design of future systems, and the analysis of existing ones with a view to finding the best possible forms and conditions of control. With this approach, many of the miscalculations and mistakes made and being made in the design and operation of power systems can be avoided.

For the most parts, these miscalculations and mistakes have had their origin in the view that the power industry is simply a system which need not be treated 'cybernetically', as an integral part of the country's national economy. Of course, the decisions taken for this 'simple' system may be correct as far as they go. Unfortunately, they do not go too far, because the power industry is a complex, large cybernetic system. Therefore, one is forced to use cybernetic methods in dealing with large power systems, and to build a 'global' model of the country's economy of which the power industry is a constituent element. This model, depicted in the form of flow and block diagrams, must reveal the deep-lying feedback paths, not always obvious at first glance, which exist between the power industry and all aspects of the country's economic and social activities, between power generation and the biosphere (ecological system), between the power industry and demographic factors, etc. The cybernetic approach is as important in the control and protection of the constituent subsystems and elements of a complex power system.

I-5. Development of Cybernetic Methods Applicable to Operation of Electric Power Systems

As a science, cybernetics deals with the general principles underlying the operation of an automated power system and its behaviour as an integral self-controlled system seeking an optimum performance as regards economy, the quality of power generated, and the quality of services rendered. Considering power-system cybernetics (Fig. I-4), it should be stressed that the design and development of machines, apparatus, controllers, protective devices, and facilities for the transmission of data and control signals

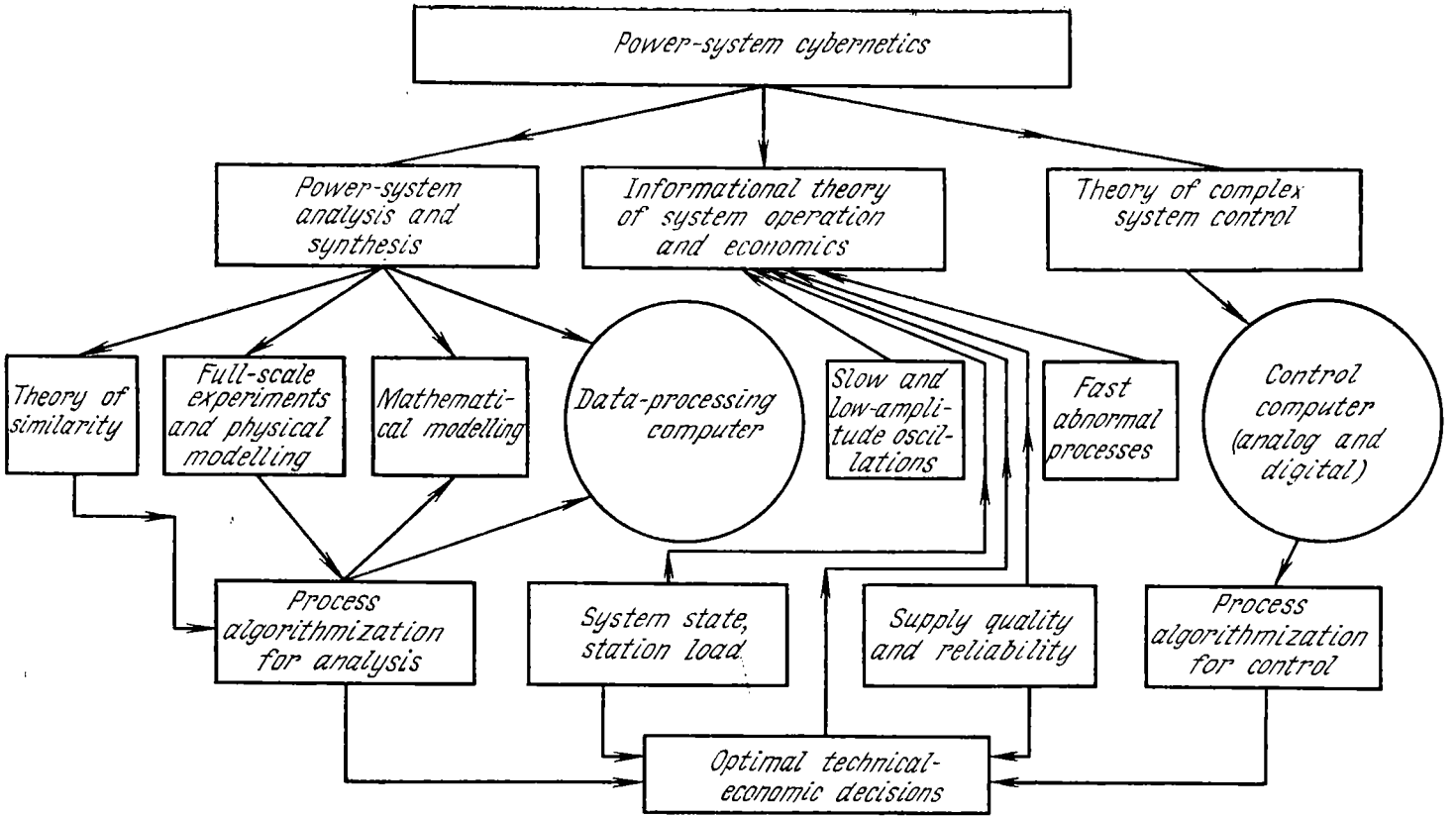

Power-system cybernetics
Power-system analysis and synthesis
Informational theory of system operation and economics
Theory of complex system control
Theory of similarity
Full-scale experiments and physical modelling
Mathemati-cal modelling
Data-processing computer
Slow and low-ampli-tude oscil-lations
Fast abnormal processes
Control computer (analog and digital)
Process algorithmization for analysis
System state, station load
Supply quality and reliability
Process algorithmization for control
Optimal technical-economic decisions

Fig. I-4

within a system lie outside its scope*. Nor is it concerned with the calculation and design of specific networks, conductors or towers for power transmission lines. Thus, despite the seemingly all-embracing definition given above, power-system cybernetics handles a limited range of matters and provides a link between the other disciplines concerned, without any attempt to supplant them.

As applied to power and, especially, electric power systems, cybernetics covers three large groups of problems, and may be thought of as consisting of three divisions.

Interaction of the constituent subsystems with one another and with other systems essential to a given power system. Here, we would need methods which would enable us to optimize system operation and expansion under conditions of incomplete or partly mutilated information. We would need a *mathematical formalism* which, utilizing practically available information, could describe optimum conditions for power-system operation in a wide sense, that is, subject to interaction with other systems. Optimization should be sought under both steady-state and transient conditions.

As a science for which the foundation was laid down by N. Wiener, cybernetics began to develop and be applied to practical uses with the advent of high-speed digital computers. However, it would be wrong to limit its role to mathematical investigations of various processes with the aid of computers. The first division of cybernetics covers the methodology of experimentation — the theory of similarity, experimental design, and modelling, which enable the investigator to test a mathematical hypothesis through purposeful experiments and to minimize the number of variables involved. The basis on which mathematical and physical models can be built from experimental data is provided by the theory of similarity, the bedrock of all experimentation theory and ordered presentation of events. As a scientific discipline, cybernetics lays down principles for the theory and practice of *physical system* and *model tests* to gain insight into the interaction of the individual system elements, of man and various automatic controls, and to see how a new component or components simulated by models will perform within an actual system. Models are especially important in adjusting the mathematical description of a complex system at work. Of course, an equal share of importance is to be attached to the theory and practice of mathematical modelling as a tool with which the interactions of automatic controlled system elements can be simulated rapidly and accurately in accordance with the hypotheses embodied in a mathematical description. Combined with physical modelling, mathematical modelling can yield complex-system characteristics and responses adequately fitting the physical processes involved, thereby supplying answers to practical problems.

In developing cybernetic control devices, we should contemplate the use of not only high-speed digital computers and elements, but also of analog

* Temnikov F. E. et al. *A Theory of Information Engineering*. "Energiya", 1971.

computers and hybrid systems essential in any control system (analog or digital), capable of adapting themselves to changing environments.

The above matters falling in the first division of cybernetics mainly relate to means refining the conditions in which the elements of a complex automatically controlled power system are to interoperate. In order that one can efficiently use methods enabling a self-adapting (cybernetically controlled) system to operate in an optimum manner, one needs data on the state of the system, that is, on the manner in which the system gathers, transmits, reduces and again transmits information. The related matters may be classed in the second division of power-system cybernetics.

Theory of information acquisition and transmission. Power systems scattered over large areas contain each a great number of interacting elements associated with a multitude of variables. This inevitably limits the capabilities of data acquisition and transmission. Therefore, in the second division of power-system cybernetics, emphasis should mainly be placed on defining the minimum information needed, coding and transmission of this information, and information capacity and reliability (noise immunity) of transmission channels*, as functions of the amount of information carried by messages and the probability of messages (see Chap. 4).

The terms of reference of this division should also include the effect of random disturbances on the operation of power systems. The point is that every power system is subjected to continuous perturbations, that is, minute changes in voltage, frequency, load demand, or power flow between subsystems. This 'breathing' of a power system is decisive in establishing margins of stability for automatically controlled systems and appropriate controls. Apart from an insight into how the system is 'breathing', it is essential to define typical groups into which information about the perturbations can be divided, the minimum required information that should be made available to automatic controls, the data important for the economic sharing of load among the stations of the system, and the minimum data about the system state (load, state variables of stations, networks, etc.) from which one can readily determine the power (both active and reactive) delivered by each station of the system, while minimizing the overall cost of the energy produced, maintaining the quality of services to consumers, and satisfying all other constraints (as regards effects on the biosphere, etc.).

No cybernetic approach to the control of system state can be realized, unless we know what minimum amount of information is needed, what quality it should have, how it is to be transmitted, and how the control signals generated by cybernetic devices are to be received and passed on to servo mechanisms at the stations and substations.

Similar problems arise in connection with methods for voltage and frequency control in power systems and interconnections. Here, too, all ele-

* Power-system cybernetics does not concern itself with the design and construction of transmission channels.

ments contributing to the common process must supply information which is to be processed by cybernetic devices, and these should pass control signals on to appropriate servo mechanisms.

Appropriate control under abnormal and fault conditions is equally important to the reliable operation of a power system. Again, a minimum information is needed on the basis of which control devices could not only respond to what has already happened in the system, but also anticipate what else can happen.

Cybernetic features have already been incorporated in some of existing controllers, for example, field controllers installed at a number of stations. These controllers respond both to changes and to the rate of change in state variables.

The matters related to establishing the minimum of essential information, its acquisition and transmission, which are cardinal problems in power-system cybernetics in general, gain especial importance in system control under transient conditions, when everything occurs at a fast rate and where any mutilation in the data transmitted is particularly hazardous.

Many examples might be quoted to show the importance of gaining and evaluating relevant information in power systems, but it will suffice, for the time being, to say that the approach to the matter is the same as in the other fields of cybernetics.

Cybernetic theory of automatically controlled power systems. This division deals with control theory adapted to the capabilities and tasks of cybernetics. Here, use should be made of the prospective advantages offered by control computers which can process input data and supply the relevant instructions to attending personnel or operate appropriate controls directly.

In this division attention is concentrated above all on the block-diagram synthesis of elements and systems incorporating analog or digital control computers which operate by continually monitoring the system state and causing the relevant controllers to adapt themselves to changing conditions.

I-6. Tasks and Courses of Action of Power-System Cybernetics

From the above listing, far from being complete, of what concerns power-system cybernetics it is seen that the topmost tasks are the development of new cybernetic control devices and systems and also of new design and calculation techniques that would adequately fit them. Work in this field, both theoretical and practical, should be focused on finding methods to influence the behaviour of a system in the desired direction and on evaluating the quantitative results of this action, rather than on a description of system behaviour, as was the case in the past. In such systems analysis, undertaken either to test the system behaviour or to solve control problems, the pivotal point must be a continuous corrective action based on statistical data, often incomplete or even lacking in credibility, about

the run of a process, rather than the deterministic accuracy of calculations based on design data. It is this aspect that is characteristic of the cybernetic approach which achieves control through interaction with the environment and by feedback. It follows then that control takes on a decisive importance in both the design and operation of a present-day power system. Lack of accuracy in the analysis of a process amounts to making extra investments in the system. To state this differently, when properly applied, cybernetic methods of power-system programming and control will save money or help generate an additional quantity of energy.

Tasks of top priority. The tasks defined above point to the need for an adequate mathematical description of a complex cybernetically controlled power system. In this connection, it is important to remember that, firstly, such a system is extremely complex and has a practically unlimited number of degrees of freedom and internal connections; and secondly, many of the processes taking place in such a system are probabilistic in nature. To simplify the mathematical description of the system so that it can be used in tackling practical and engineering problems, it is necessary to get rid of the degrees of freedom and connections superfluous to a given class of engineering problems and only to retain those essential to the representation of the system's properties.

A completely novel point in the mathematical description and block-diagram synthesis of a power system is that the control devices are incorporated in these diagrams as a single entity.

In treating a complex system as a cybernetic one, we need a novel approach to its analysis, which would combine studies into the system behaviour and the choice of control techniques. Novel techniques must be used in synthesis, too, leading to block diagrams that would maintain the process on a predetermined course in a system where state variables change in a random manner and where these changes actuate the controls so as to guide the process as required. Of course, such an analysis must show whether the process can be maintained with existing physical plant or there is a need for some other facilities. Thus, changes are needed in the very approach to design which must consider processes as controlled and subjected to continuous corrective action in a desired direction.

Among the immediate tasks of power-system cybernetics is the development of a synthesis method and, on its basis, of diagrams that would incorporate multi-variable controllers acting simultaneously on the excitation system and the prime movers, and also devices which would improve the system state and change their own settings according not only to the present state of the system, but also to the state anticipated by the analysers (state estimators) of cybernetic controllers.

Another important task facing power-system cybernetics is to develop cybernetic devices that would control load sharing between the system stations. These devices must be supplied information about the current state of the system elements, process it, and generate control signals for appropriate servo mechanisms built into the generating units of the stations

and the switching centres of the system so as to optimize the economy and other performance criteria (for example, reliability, stability, etc.).

The ultimate goal, the completely optimal cybernetic control of a power system, cannot however be achieved without some intermediate stages in the development of control methods.

Methods of analysis and synthesis applicable to power systems. Very often the methods and techniques used currently in power-system studies and system-control development are loosely called analysis or synthesis.

As will be recalled, *analysis* refers to a method of investigation by which an entity is resolved into its constituent elements for the purpose of studying their properties. This method is used in all sciences where the object or problem of interest is dissected in the mind's eye by abstract logical reasoning.

Synthesis refers to a technique of investigation by which individual elements are combined into an entity to reveal what new properties or qualities this union can provide.

Synthesis is inseparable from analysis and is an essential method in the cognition of reality. It is only in the most elementary cases that synthesis reduces to simply adding together existing pieces of knowledge.

Scientific thinking usually investigates objects that can only be viewed as natural entities not reducible to a simple sum of their components. In the cognition of reality, synthesis establishes relations between the various aspects and the various elements of the entity under study. Synthesis is impossible and inconceivable without analysis, that is, without resolving the entity into its constituents. Analysis and synthesis form a dialectic union which is realized in every logical operation. In a specific investigation, however, these operations of thinking may occur in turn.

The concept of synthesis is widely used in the various divisions of science, technology and art. In chemistry, there is organic synthesis by which organic compounds are derived from the simpler components. In mechanics, the synthesis of mechanisms has as its objective to derive mechanisms which posses specified kinematic and dynamic properties. In kinematic synthesis, the type of mechanism is assumed to be specified in advance, and one seeks to find for its links the dimensions that would enable them to have the desired positions, velocities and accelerations. In the field of art, synthesis produces an organic combination of various arts in a single work of art, say graphic and decorative, which creates an integral artistic impression.

The large-scale use of automatic control in electric power engineering calls for novel techniques in system analysis and synthesis. Of late, linear graph theory has proved fruitful in the analysis and synthesis of electric networks. Some elements of this theory have in fact been in use for a number of years. This theory, worth of a very careful attention, must become an integral part of power-system cybernetics.

The fundamental task in the analysis of an electric system is to establish mathematical relationships between the characteristics of the system and

of its constituent elements according to the manner in which these elements are connected. The task of synthesis is to find at least one set of characteristics and system elements and a connection scheme for these elements that would yield the desired set of system characteristics with a tolerable degree of approximation.

What is common to all known methods of synthesis is that the choice of components and of their connection scheme is mainly based on the past experience gained from systems having similar properties. In this sense, the first stage in the design or development of a system is more of an art, an art also in the sense that whether or not the chosen components can be connected in the chosen scheme depends on the physical realizability of the resultant system.

The design of a system starts from an idea; for its realization methods of analysis, although necessary, are not sufficient. A working knowledge of existing system elements and of their characteristics gathered from past experience is of course valuable, but it cannot guarantee that the final result will be a correct one. The system designer may be said to be in the position of a composer or a poet. The choice of sounds, accords and tunes, or of words, phrases and expressions that may be composed into a piece of music or a poem is limited only by imagination. Of course, the composition must be in agreement with certain rules, or else the result would be simply noise or run counter to grammar. The composer follows his ideas and tests them with a musical instrument. To the engineer, the instruments with which he can test his pick of concepts are mathematical analysis and laboratory experiment. But, much as knowledge of only the theory of music or of grammar is not enough for anyone to become a composer or a poet, so knowledge of only theory or methods of analysis is not enough for anyone to be an expert in a particular field of engineering.

Stages in cybernetic control. Integration of cybernetics in power-system engineering does not imply that man must be ousted from all levels of control in a power system by 'thinking', 'decision-making' and 'executing' machines. The goal of the cybernetic approach is entirely different. Above all, it requires that a power system be treated as a complex system made up of interrelated feedback subsystems, and that the system itself be looked upon as a subsystem of a still bigger, *global* system encompassing the country's economic and social life.

If we trace the path along which the methods of control have so far developed, four major stages may be discerned in the relations between man and the control computer.

1. In the first stage, computers act only in a *consultative* capacity to the engineer who controls a system. The data gathered from the controlled system are reduced and inserted into the computer by man (or with his aid), and the computer puts out *recommendations.* These recommendations are executed by the engineer who may adjust the machine solution in the light of his past experience.

2. In the second stage, computers receive *some of the data directly from*

the controlled system through appropriate automatic transducers. There may also be provisions for the transmission of computational results and appropriate recommendations to some control elements of the system. However, the computers cannot yet feed control commands directly to the servo mechanisms concerned or to the switching centres; this is done by man.

3. In the third stage, a computer which receives all or part of the data directly from the controlled system not only generates recommendations to be executed by attending personnel, but also *feeds at least some commands* directly to servo mechanisms and switching centres. Here, man and machine work together in such a manner that man, relying on his experience and intuition, can adjust the control effected by the computer. It is not unlikely that the computer in this combination may be of a self-adaptive type, capable of modifying its control algorithm in the light of the experience it gains in cooperation with man.

4. In the fourth stage, *control is fully automatic*, which implies that the computer receives all data directly from the controlled system and automatically disseminates all control commands. This does not, however, mean that man is completely divorced from the control of the system. Simply, his functions are now reduced to keeping watch on the run of the process and making appropriate changes in the control algorithm, which amounts to a continuous expansion of the algorithm base of the system.

Thus, the cybernetic approach to the control of a power system is at the same time an *interactive approach* which leaves room for man to contribute his experience and intuition, thereby expanding the control algorithm.

The basic difficulty in making the control of power systems fully cybernetic lies not in the lack of appropriate algorithms or sufficiently speedy or reliable computers, but in the fact that the power generating plant and switching apparatus of power systems have been developed without concern for cybernetic control and fail to meet its requirements. It has been found that the servo mechanisms that control power-generating plant respond differently when they are actuated by man or by pulses from a computer. Cybernetic control requires not only specially designed servo mechanisms and feedback elements, but also specially designed power plant. Already the early successes of automatic fault-clearing and preventing devices and field controllers made it possible to improve the design of synchronous generators (to make their rotors lighter, to increase reactances, etc.). It is beyond any shadow of doubt that in the future advances in control will feed back increasingly more improvements to plant design. It is misleading to speak (as people often do) of 'automated' systems or system apparatus because this implies that they were designed as non-automatic and were automated at a later time, which is ridiculous, because power plant must be designed and automated at the same time.

Cybernetics in power-system design, prediction and performance evaluation. During the design of power systems and apparatus, cybernetics finds

use in that design work is now mostly computer-aided and also in that account is taken of the many relationships existing within a system that make it cybernetic.

In computer-aided design, use is mostly made of digital computers in combination with interactive methods, that is, a direct dialogue between man and machine with man evaluating the results as they appear and making corrections to the algorithm. In the computer-aided design of power transmission lines, for example, it has been found convenient for man and machine to converse using a visual display terminal built around a storage cathode-ray tube, and an electronic light pen. The data set into the display terminal are stored in the computer memory and regularly updated. The solutions that the computer puts out are displayed on the storage tube, and the necessary corrections to the design are inserted with the light pen. Thus, the design process is iterative, and the designer can see all the intermediate stages in his progress towards the final solution. A typical program for computer-aided design is in three parts, one part controlling the video display terminal, the other executing instructions and controlling the input device, and the third carrying out the actual calculation and optimization of the electric network being designed.

As already noted, system design and, the more so, system expansion forecasting must be based on a consideration of the system as a complex entity made up of subsystems and interacting with other systems. In this connection, it is clearly seen that cybernetic methods cannot be reduced to a simple shift of some computational work on to digital computers*. The use of cybernetic methods implies more than that, namely fundamental changes in the methodology of technical and economic studies, a novel approach to these studies so as to bring out and assess the ramified internal relationships and feedback which remain hidden or unnoticed in the conventional approach to the design of new and the analysis of existing systems. Cybernetics must seek novel technical and economic decisions supported by more accurate and realistic data, such as cost and cost formation, price, profit, etc., that must be supplied by political economy and other fundamental sciences. But even before these complex economic matters are clarified, the results of analysis can be made more reliable, if power systems are designed or predicted in the light of the factors that make such systems cybernetic. At present, the design of power systems is still suffering from lack of attention to competitive designs which are either left out of consideration altogether or examined formally, without proper evaluation, or still worse, discussed often from purely emotional considerations. This forcefully bears out the fact that calculations made by a computer cannot, as such, guarantee a truly objective judgement and also that technical and, the more so, technico-economic calculations would inevitably be biassed

* In principle, cybernetic methods might be visualized to be used without digital computers. However, they would not be properly developed and applicable to practice, and they would be doomed to the same fate as Ampere's ideas.

if they ignore competitive designs and if they leave out of consideration the dynamics and interaction of systems with the national economy.

Thus, a correct decision can only be obtained with a fundamentally cybernetic approach, that is, an approach to a system as a large and complex (or global) organism with all of its ramified internal relationships, and not by simply using computers. However, this cybernetic approach must be supported by a general economic investigation based on appropriate hypotheses of evolution and on models of a global type from which it would be possible to assess various system expansion and control alternatives in relation to, rather than in isolation from, the entire national economy and to predict how it would be affected in each case. Without such *evolutionary models** giving a true representation of feedback between systems and subsystems, economic mathematics in general and economic cybernetics in particular cannot be effective. Individual systems (industries or sectors) and subsystems seek their own goals for which one can find optimal conditions; however, finding these optimal conditions does not necessarily lead to the optimization of the entire system. True optimization is only possible on the basis of global models which take account of all relationships between all systems and subsystems, that is, on the basis of a cybernetic approach. Of course, this approach involves the use of computers. Without them, the time required would be prohibitive.

Cybernetic methods must likewise be used in the design of power-system elements. Here, too, the designer faces a multitude of problems, such as the realization of a reliable system from unreliable elements, evaluation of, and improvements in, reliability, etc.**

In some of the above cases, it will be necessary to proceed from the fact that the system behaviour is probabilistic and that more than one equally probable solution may be obtained. In the circumstances, optimization at the design or prediction stage and its analysis as regards complex power systems are extremely complicated, and it is vitally important that any theory be followed up with thorough experimental verification. It may be noted that the application of 'ready-made' mathematical methods, although extremely promising theoretically, has not been always effective. For example, dynamic programming as it exists today can only be used to formulate, but not to solve problems related to power systems and power-system engineering.

It is still unclear how a number of new optimization techniques worked out by mathematicians can be applied to power systems. It is not unlike that, as regards power and some other multi-loop, multivariable systems,

---

* Makarova, A. S. et al. *Optimization of Electric Power Systems by Use of Mathematical Models.* USSR Academy of Sciences Press. "Energetika i transport", 1966, No. 5.

** An example is offered by bionic modelling where electric network models simulate systems existing in animate nature. Use may be made of the analogy with a reliable network formed of unreliable neurons, or redundant systems as exemplified by the self-redundant blood circulation systems of some animals.

it would be reasonable to seek a sub-optimal, and not optimal control. Indeed, if instead of a rigorous performance criterion, there is only a hypothesis as to what is good for a power system, it would be futile to try to find an optimum solution. Instead, the goal must be a definite system expansion and control strategy found by means of cybernetics.

In the study of an expanding power system, resort is sometimes made to the concept of system-state entropy. This concept may be used as a measure of reliability in system design and also in the design of automatic fault-prevention means, in the evaluation of the complexity of system control, etc. Work done to date in this respect has shown that the design of control means for an expanding system under fault conditions, which is a very complex problem, can be markedly facilitated through the use of cybernetic concepts and methods, too. However, cybernetic methods in general, and those of power-system cybernetics are not something entirely new or built from scratch. Cybernetics has grown out of the findings and knowledge accumulated by other sciences and from the evaluation of computer capabilities. Power-system cybernetics has stemmed from the approach to the power industry as a large, complex, expanding system. In fact, foundation for this cybernetic approach was laid down in the GOELRO plan (see Sec. I-1) more than 50 years ago. Of course, that plan did not, and could not, use the present-day terminology such as 'large system', or 'cybernetic system'. This, however, does not diminish the value of the GOELRO as an integrated plan encompassing the many subsystems that make up the country's national economy. Thus, the concept of power-system cybernetics is a logical extension of the earlier theoretical concepts.

I-7. Optimization of Large Power System Expansion and Operation*

As regards power-system management and control, the overall problem of optimization can be segregated by a well-defined delineation into a range of tasks, each associated with a definite span of time, as follows**: (1) long-term forecasting of system expansion for 20 to 30 years ahead; (2) short-term forecasting or prospective planning of system expansion for 10 to 20 years ahead; (3) prospective planning of system expansion for a period of 5 to 10 years; (4) economic dispatching of system expansion for 1 to 2 years; and (5) operational system control over a year. These time stages differ in tasks and goals, credibility of input data and, as a consequence, in the methods used to handle sometimes the same tasks.

The temporal stages need not be rigidly tied to definite dates. Power-system expansion control and management must be flexible. However, the

---

* See Astakhov Yu. N. et al. *Characteristics of, and Interrelations between, the Tasks of Control of the Integrated Electric Power Supergrid of the USSR.* "Energetika i transport", 1970, No. 6.

** This is not a universally adopted classification. Sometimes, the division is into three forms of forecasting (long-term, medium-term and short-term), or into forecasting and planning.

stages related to the planning periods of the national economy must be singled out.

Long-term forecasting is concerned with defining the most promising basic directions in system expansion and in the technological progress as regards the development and construction of basic system elements. Short-term forecasting deals with the structure of system expansion and with the principles and methods of system state control. Long-term planning has as its task to draw up a long-term plan for the development of power-generating capacity, to define the configuration for the main networks and to select system control means. The task for economic dispatch is a yearly plan for system expansion and operation control, and that of operational control a short-term plan for system control and actual routine control of system operation.

The tasks in the various temporal stages are related to one another by the requirement that system expansion be continuous and correlated. In fact, matters related to system control must be settled already at the time of long-term forecasting of system expansion.

The conditions under which system operation will be effected cannot but bring pressure to bear on the choice of station type, network structure, and network handling capacity. Among other things, the power generating capacities of a system must be sited so as to facilitate control of system operation under normal conditions, to supply the peak load and to provide enough reserve capacity needed to make up for any emergency drop in generation in any part of the system.

The primary-network structure must be such as to permit the segregation of the system into zones or areas. Such segregation may be necessary in an emergency so as to localize a fault or in anticipation of a fault (segregation at the nodes of the system's 'weak' links).

In formulating the tasks to be achieved during the various time spans, it is important to bear in mind that:

(1) the optimization of expansion for any electric power system is directly related to that of the country's pan-energy system (fuel and energy supplies) at both the national and the regional levels. This relation is mainly one of the quantity, forms and rates of power generation and utilization, including heat supplied by district heat and power stations, during the various seasons of the year and in the various regions of the country. It is to be noted that this relation can be evaluated to a varying degree of accuracy, depending on how detailed the relevant studies have been;

(2) the problems that make up the overall task of power system control belong to fundamentally different levels. They may be divided into those concerned with the choice of the system structure, system state or operating conditions, and the parameters of the system elements. The structure problems are solved as part of the optimization of the pan-energy and electric power system for a period of 10 to 20 years. Among them are the choice of types, capacity, sites and construction priorities for the main power stations and transmission lines. The solution of system-state problems

yields the optimal, limiting and abnormal operating conditions. Within the framework of short-term forecasting these problems are of auxiliary nature, only supplementing the decisions taken in handling structure problems. In contrast, they come to the fore within economic dispatch and operational control of the system. The selection of optimum performance indicators for the individual system elements is related to system control only indirectly. The matters involved here can be tackled, assuming that their relations to the system are unique or may have alternatives. As a rule, in the control of system expansion, there will be a multitude of valid state and system variables which may be taken as realizable in a number of alternative designs;

(3) the scope and degree of detail in the various tasks vary with time and from area to area. In the study of an electric power system as a whole, it is assumed in advance that one must give at least an approximate consideration, over the same time span, to its constituent elements (interconnected and regional power systems and individual installations). Their subsequent study has as its goal to refine the solutions derived in the previous investigation and is conducted with account of the feedback relations disclosed.

Thus, electric power system control involves a sequential (often independent) solution of the individual problems and subsequent refinement of the decisions made. This is usually done both in a centralized way and locally, by many organizations. To do this properly, these organizations need a system of quantitative criteria leading to solutions that would be optimal for at least the system and, ideally, for the national economy as a whole, even though the solutions are sought locally. Sometimes, this purpose is served by what is known the 'closing costs' of energy resources and electric power.

The basic design data whose credibility differs markedly at the various time levels of problem-solving are: the rate of technological advance in the power and allied industries, demand for electric and thermal energy, the load curves of the system's nodes, the power resources available and the likely rate of expansion in their utilization, the technical-economic and performance characteristics of the power stations, transmission lines, and automatic control means (including data on system security), and the lead time needed for construction of power installations.

The fact that these data lack in credibility or essential detail inevitably makes the tasks involved probabilistic in character and necessitates a choice from a multitude of alternative solutions. It is important to stress that the acquisition of design data likewise poses a problem within the subject-matter considered.

Long-term system expansion forecasting. This stage is mainly effected within the framework of the pan-energy system (while optimizing the fuel and energy balance). It seeks to define objective trends in the expansion and technological advances of the power industry, the most likely changes in demand for electricity and heat, and also for the staple forms of power resources, with different forecasts for their availability.

In long-term system expansion forecasting organically related to long-term forecasting for the pan-energy system, the primary task is to form estimates of: (1) the likely rates of technological progress in electric power generation by existing and new thermodynamic cycles, with different forecasts for the availability of various forms of energy resources; (2) the likely rates of technological progress in power transmission; (3) the resources (including capital investments, basic plant and materials, utilities, manpower and annual expenses) needed to maintain the adopted rate of expansion in the generation and transmission of electric power.

The above estimates serve as a basis for defining the range of major scientific problems, economic studies, research and development work, and experimentation needed to turn the forecasts into reality.

For a plausible estimate of what the power industry would be like in the year 2000, we would need substantiated forecasts for: (a) the expansion and technological progress of nuclear power generation, including heat generation; (b) the status of MHD and steam-vapour power plants using gaseous, solid and nuclear fuels; (c) d.c. power transmission; (d) a.c. power transmission at 1000-1800 kV; (e) power transmission using superconductivity and waveguides; and (f) the prospects of a wider use of electric power for traction and propulsion.

Long-term forecasts should mainly use expert opinions based on both objective quantitative trends in the expansion of the country's economy and power industry, and on intuition.

The foregoing applies to forecasts for 25 to 30 years into the future. As important are detailed forecasts covering periods from 15 to 20 years in advance, as they may throw much light on the trends brought out by short-term forecasts. As it appears, this form of forecasting should use basically the same techniques as short-term forecasts, but with markedly less attention to detail.

Short-term system expansion forecasting. This is the main stage as regards the solution of structure problems and the selection of control strategies for an electric power system, as it spans the same time interval which passes between making a decision and bringing the basic power units to operation at full capacity. This is why this stage should include system studies at the various geographical levels.

The first level is concerned with the ultimate structure of the electric power system as part of optimization of the pan-energy system. This optimization seeks to determine: (1) general solutions for the expansion of fuel and power supplies as an element of the national economy; (2) optimal proportions for the expansion of the electric power, gas, oil and coal supplying integrated systems.

As the final result, this optimization outlines the domain encompassing system expansion alternatives with the highest probability of optimality, each stating how much generating capacity should be put in operation at the various types of power stations in each region of the country, the forms of power resources to be utilized by these stations, the function and capa-

city of interregional power transmission lines. At this stage, the regional and interconnected electric power systems are considered in a general outline.

The second level is concerned with the optimization of expansion for the electric power system proper. This might be called a 'parochial' approach, but it is justifiable after we have established and optimized, in outline form, the external flows and connections that allow the country's electric power system to be isolated from the pan-energy system for a more detailed investigation. These external flows and connections are the total production of electric power over the country and an approximate breakdown of electric power commitments among the various systems, an approximate withdrawal of heat from district heat and power stations, the quantity of power resources allocated to the electric power stations of individual systems, and their approximate seasonal availability.

The overall task in the analysis of the electric power system as an individual system for 10 to 20 years into the future may be broken down into the following sub-tasks:

(1) the choice of the basic scheme for the electric power system (d.c. voltage levels, a.c. voltage levels above 500 kV, interregional d.c. tie-lines, etc.), the choice of a hierarchy of electric connections with consideration of power stations rated at 1500 MW and more;

(2) the estimation of likely normal and basic inter-system abnormal states so as to make the necessary refinements in the system structure expansion adopted earlier;

(3) the choice of policies and methods for the operational control of the electric power system and for the acquisition, processing and transmission of data essential to the operational control and management of the electric power system (exchange of planning and economic data);

(4) the search for the material resources needed for the expansion of the electric power system over the period in question;

(5) the formulation of specifications for basic plant, operational control means, and information-exchange facilities.

The analysis of the electric power system at the second level calls for more detailed data on electric power commitments and dispatch among the constituent systems, the likely contributions of district heat and power stations, and also on the heat and electric power commitments and dispatch of district heat and power plants. It is important to remember that the optimization of the pan-energy system is related to, and affects that of, the constituent systems in a very complicated manner and to such an extent that in some cases it may be necessary to draw up and evaluate parallel alternative designs for them.

The third level is that of a detailed expansion analysis of individual interconnections and regional electric power systems. This form of analysis is essential where the need arises to correlate and coordinate the expansion of regional power systems after detailed data on demands for electric power become available. Here the degree of detail is fuller than it is in the ana-

lysis of the country-wide electric power system. Among other things, more detail is filled in as regards local tasks such as expansion plans for centralized heat supply, district heat and power plants, the energy resources used, etc.

It is at this level that an optimum structure should be defined for the individual regional systems, including the connection scheme and characteristics of the primary network, the siting, characteristics and basic plant mix of major power stations, spinning reserve schedules, and forms of system control under normal and abnormal conditions. The material resources needed for system expansion should be defined at this level, too.

It is to be noted that the detailed optimization of regional systems may sometimes necessitate refinements in the decisions made earlier for the electric power system as a whole. This is why particular care must be exercised in establishing and evaluating the external flows and connections between the electric power system and the regional general power systems.

A short-term expansion forecast leads to the formulation of design specifications for the individual installations that make up the electric power system.

Long-term system expansion planning. At this stage, emphasis is placed on the analysis of interconnected and regional electric power systems as economic entities. The objective of long-term planning as regards regional systems is the refinement of: (1) the power resources allocated to a given system; (2) the due dates for the design, construction and commissioning of the system's major installation; (3) balances of electric and thermal power, annual and typical day's load curves; (4) the requisite handling capacity of the various power transmission lines; (5) the system configuration and parameters adopted for 5 to 10 years into the future at the previous stage, the security indicators, and the siting of spinning reserve; (6) matters related to the operational control of the system under normal and abnormal conditions and to the design and local management information networks; (7) the quantities of all material resources needed for system expansion.

The principal design work that constitutes the basis of a five-year plan for system development includes expansion designs for: (a) the system transmission networks; (b) the distribution networks in the metropolitan and rural areas served by the system; (c) the heat supply systems of the metropolitan communities within the service area of the system.

Thus, this stage produces data for inclusion in the five- and one-year plans for the construction of power installations. The design data used at this stage are far more credible and the system can therefore be analysed in greater detail.

Economic-dispatch control of system expansion for one to two years. This stage involves decision making at several geographical levels (for different power installations), although the tasks are formulated in about the same manner. The degree of detail in system analysis at the various geographical levels varies according to the degree of equivalence used. The principal

tasks of economic-dispatch control are: (1) to predict variations in electric and thermal loads for the system regions (centres) over different spans of time throughout a year; (2) to develop annual discharge and inflow schedules for hydro-station reservoirs and for cascaded hydropower stations; (3) to dispatch reserve capacity and select a connection scheme for the main network from security analysis and the optimum system states under normal conditions; (4) to develop a generating plant maintenance schedule for the year; (5) to formulate recommendations for the use of control means for the year; (6) to refine priorities for commissioning system elements and for the construction of new installations.

This stage turns out a yearly plan for the system expansion and operation, which includes dates to start the construction of new system installations. Where the starting dates are to be fixed for major projects (such as a large power station or a long-distance transmission line operating at a super-high voltage) with a lead time of three to five years, the system should be analysed in the light of the next five-year period, because the five-year plan already approved cannot cover so remote a perspective.

Short-term operational system control. This stage is concerned with system control within days, weeks and months of a year. This task mainly reduces to refining and making more specific the recommendations based on the decisions taken in tackling the first five tasks of the previous time stage. A distinction of this stage is that it is concerned with the routine control of the system under normal and abnormal conditions as a task in its own right. This task must be coped with at the pace of the system's operation. This is why a good deal of importance at this stage is attached to the many computations (steady-state conditions, stability, short-circuit currents) involved in the normal, scheduled outage and other specific operating states of the system.

CHAPTER 1

EXPANSION FORECASTING

1.1. Economic Analysis and Tasks of Forecasting

Economics has traversed a long path of creative evolution. In fact, it is today a system of sciences. In the USSR, the core of this system is the political economy of socialism, which gives the widest theoretical generalization of the historic experience in the planned construction of a new society in this and other socialist countries. Political economy has provided a theoretical basis for specific economic sciences which organically combine theoretical analysis with the solution of applied problems in the national economy as a whole and in the individual branches and industries. It is typical that in one way or another economic problems have now become of concern not only to economists, but also to technologists in all branches of the national economy and in all fields of engineering. This is why engineers concerned with electric power systems in general and with power-system cybernetics in particular must have a clear idea of forecasting and of the techniques it uses.

Forecasting is a complex and multi-faceted problem because the development of present-day productive forces is directly related to advances in the natural and technical sciences. In socialist society, an increasingly active and reforming role is played by the social sciences. This is particularly true of economics which deals with the economic basis of society and the evolution of socialist production relations in direct interaction with the growth of productive forces. Because of this, economics is called upon in all of its studies and investigations to contribute actively to the union of the huge potentialities of the scientific and technical revolution that is currently under way, with the capabilities of socialism. This is the crucial point in the building of communism and in the economic competition of socialism and capitalism. It is vitally important to remember that progress in technology, however important it may be, is not the goal in itself. Even the efficiency of an industrial installation does not tell a whole story. The 'whole story' can only be narrated when one resorts to economics. Today, technology can cope with a host of mammoth tasks: it can dam seas, divert rivers, and so on. However, before any project is to be launched, one question must be answered: How much will it cost? What will be the pay-off? Because of this, one has to look into the future, that is, to resort to forecasting and planning. It is logical, therefore, that the 24th Party

Congress dwelt at length on the matters of forecasting and planning. As is stressed in the Report of the Central Committee of the CPSU, all successes in the socialist economy are inseparable from planned management. The quality of planning will have a decisive effect on its future achievements. This is why emphasis should, as before, be placed on its streamlining and perfection.

The need for higher scientific standards in planning is dictated by the conditions in which the country's economy is developing and by the fact that the Soviet Union has reached new proportions in its economic development and has built a powerful economic potential. Advances in science and technology and the growing needs of society have been accelerating changes in the field of material production, in the sphere of circulation, and in the service industries. There has been a continuous expansion in the range and choice of manufactured products, a progressive cut-down in the time necessary for novel products to be put on the market, a steady expansion of interrelations between the various industries, and a growing complexity in production re-tooling and set-up. All of these factors necessitate rapid application of scientific and technological advances to industry and add more weight to the role that the state can play in the field of structural changes important to meeting the vital economic and social objectives. In the circumstances, it becomes mandatory to assess all alternatives of the development programs at regional and branch (sector) levels.

It is particularly important to gain deep, concrete and all-round insight into the laws that govern the creation of a material and technical basis for communism, its qualitative and quantitative criteria, and the accompanying processes. To quote V. I. Lenin, communism is inconceivable without technology based on latest advances in science, without improvements in labour productivity. This is especially true now that we have reached a position where we must map out a general direction for the technological development of socialist production and formulate a long-term strategy for the integrated mechanization and automation of production, while concentrating on the creation of a material and technical basis for communism. Because of the huge investments involved, it is important to have a sufficiently precise idea of what the productive apparatus of future society should be, how the advantages of a planned economy can best be utilized in the construction of communist economy. This must not only be achieved, but achieved in the best, that is most economical, way.

Technological forecasting and planning. In the climate of the new scientific and technological revolution, when big strides are being made in all divisions of science and technology, especially in branches related to the power industry, any plans and designs must of necessity take into account how a given division of engineering and all related industries will be developing in the future. Of course, much of this future development depends on advances in science. This is why early forecasting is needed in order to form a proper estimate of the measures that are to be taken to expand the power industry, to earmark the research and development projects to

be speeded up, and to select preferred methods for the generation and transmission of energy. This need is also underscored by the fact that in the conditions of the scientific and technological revolution all industry, including the power industry, is making big strides unseen ever before in man's history. Mistakes in the use of scientific and technical information and miscalculations in the estimates of future power systems might, and in fact do, have unfortunate consequences, such as

(1) an incorrect estimate of technological progress and of the resultant capabilities of industry and science, because in the conditions of the scientific and technological revolution science merges with production. As a result, the resources available might not be utilized in the best way and the pay-off would be anything but optimal. It may be added that this situation is a fairly common occurrence;

(2) adverse effects of industrial production on the environments with irrepairable or nearly irrepairable damage to the biosphere;

(3) neglect of research and development in technology and failure to maintain it at an appropriate level, with the resultant loss of precedence on an international scale and adverse effects on defence and policy. If it had been late with fundamental and applied research in the field of nuclear fission, the Soviet Union would have had no A-bomb for defence in time;

(4) a wrong sequence in the construction of projects. As an example, we may imagine a hydropower station built before energy-consuming factories (say, aluminium works) and suitable power transmission lines are completed. Although there may be low-voltage lines to carry power to distant consumers from the new hydro station, and although this hydro may produce electric power at low cost, the overall cost of electricity from such a system would rise rather than fall, because the transmission of power over low-voltage lines entails heavy losses. Also, some of the capital invested in the project would remain in effect 'dead' because the pay-off is deferred. This might not have happened if it had been invested elsewhere. As another example, picture a low-voltage transmission line built in some locality. Before long it turns out that the line's capacity is inadequate to cope with the growing demand. Here, too, the planners have been wrong in their forecasts of the region's development and in their choice of the voltage level for the line. In such cases, resort may be made to the scheme widely used at one time in France, where towers for power transmission lines would be those rated for 400 kV, but the conductors would be chosen temporarily to carry 220 kV. With this arrangement, no extra money would be spent on towers, should the need arise for a line of a larger capacity.

As still another example, suppose a plan is afoot to build a dam. This will certainly change the river's runoff and might lead to a rise in groundwater level and, as a consequence to the salination of fertile land. Therefore, the water-development plan should give proper consideration to land utilization; otherwise, an irrepairable damage might occur.

If a large fuel-fired power station or a factory using big steam-generating plant is situated,the prevailing winds cause ash and other pollutants

to lodge in the nearby built-up area. It may (and actually does) so happen that people will be forced to leave town for the country and commute daily to work.

Such examples could be multiplied.

Forecasting is a variety of management which invokes the ability to look into the future. In turn, industrial management in general and power-system management in particular is becoming a science, and a vitally important science at that, because a good deal of matters hinge on the dependability of its findings.

Both management and forecasting should proceed so as to avoid lop-sidedness and bias, to see everything in a true perspective, in the actual train of cause and effect, and to take account of the laws governing all developments.

Soviet planners who profess Marxist-Leninist views apply the general principles of dialectic and historical materialism to problems of forecasting in a creative manner. This Marxist-Leninist approach to social life requires that any factors be examined from all sides, in all relationships, in all manifestations, however contradictory, and that material situations and circumstances be unravelled from all view-points. Inquiry into the future is a vital element of man's cognition of nature and society.

In technological forecasting, too, it is important to look into social processes, to see what course they may take, and to foresee how technological processes can be harmonized with the environment. The world around us is dialectical, and an approach from all likely angles is essential. Any process involved here is anything but static — it is fluid and changing all the time, much as social processes are, and any process in the making carries in itself the remnants of the past and the seeds of the future. This is why, in probing into any object, one has to answer what this object was, what it is, and what it will be. The 'was', 'is' and 'will be' merge together in a single analysis used in forecasting. From a general, philosophical point of view, prediction is an indispensable link in the chain of scientific and technical cognition, a link stemming from the dynamics of objective reality and from the objective and dialectic reflection of this reality in man's conscience.

Man and human society as a whole, comes to inquire about the shape of things to come at a sufficiently high level of development. In principle, this is a major function of any study. A scientific study is the ability to foresee the course of a process. In this sense, forecasting is a highly advanced form of scientific activity. At present, forecasting — scientific foresight — is becoming a must in many divisions of man's endeavour. As a scientific school of thought, it did not come into being with the advent of 'future research', as some scholars insist. A consistent theory explaining why and how human society develops dates back to the works of K. Marx, F. Engels and V. I. Lenin. History has borne out how sound the founders of Marxism-Leninism were in their general predictions which could not of course cover minor technical details. Yet, even in the narrower sense — with

regard to the prospects of scientific-technical progress and economic development, V. I. Lenin's *Draft Plan for Scientific-Technical Work* and the GOELRO plan drawn up on Lenin's initiative and with his participation were the early masterpieces of scientific forecasting and provided the basis for state plans for the first time in history.

The forecasting of scientific-technical progress and economic development is a study into their future. It is of especially paramount importance to the Soviet Union. As A. N. Kosygin stressed in his discourse at the State Planning Committee (GOSPLAN) in March, 1965*, "We must have scientific forecasts for development in each industry in order to pave the way for all things advanced and progressive, in order to know in which direction plans should be developed". Forecasting for a particular industry, notably the long-term forecast (up to the year 2000) for the power industry of special interest to us, can only be developed on the basis of a more general forecast, drawn up in several alternatives, for the country's entire economy.

Special-purpose forecasts abroad are drawn up by a great number of firms. For the most part, they cover highly specialized subjects and serve industrial corporations, although there are large institutions and committees, mainly at academic and government agencies. Some turn out forecasts covering a wide range of sociological problems, natural resources, science and technology.

Management and forecasting, that is management in its wider sense in general and as applied to the power industry in particular, should be treated from two angles, geographical and time.

In the *geographical approach*, the power industry, or more correctly, the country's power system as shown graphically on paper, is analysed within the various areas of the country's territory. It may be decomposed to show separately a station, a transmission line, subsystems, and interconnections. Separate consideration should be given to mines, fuel-supply lines, etc. By dissecting the overall power system across the country, we obtain what may be called a geographical section of the system. Of course, an interconnection should be treated differently than a separate station, although there will be much in common, because the objective is the same in all cases.

In the *time approach*, system control and management is viewed differently. The *first time aspect* is the forecasting of power-system expansion. The crucial question here, to which no unique answer exists, is a forecast into the remote future (25 to 40 years). An expansion forecast for the power industry needs above all an estimate of power demand over the time span of interest. This immediately prompts another question as to how this needed power is to be produced. Will it pay, in the future, to abandon hydro or thermal plants and to go over completely to nuclear

* See also Kosygin A. N. *Improvements in Scientific Substantiation of Plans — the Crucial Task of Planning Bodies.* "Planovoye Khozyaystvo", 1955, No. 4.

power or MHD-generators? Or will their place be taken over by fuel cells? Depending on the answer, money will be put into research and development work in the respective field. An expansion forecast should analyse trends in the use of power in industry, agriculture, transport and in the homes. For example, in the United States of the 1950s no residences used electric heating. Today, the homes using electrical heaters run into tens of millions. Because of an erroneous forecast, this development came as a surprise for the US power utilities, and they had a lot of problems to tackle, especially in supplying the peak loads due to electric heating. At the opposite extreme, it may so happen that a new technology will cut down rather than raise the demand for electricity, as might be expected. Forecasts are likewise important as regards methods for long-distance power transmission. Should we gear power transmission lines to direct current, or lasers could be a better choice? All of these questions are part of electric-system expansion forecasts.

Instead of 25 to 30 years, expansion estimates may cover intervals of 50 or even 100 years. Still, scientific forecasts must not degrade to guess, but remain analyses based on sound methods. Quite a number of forecasting techniques have been developed to date. In all of them, the pivotal point is an analysis which encompasses not only the present status of science and technology and their likely progress in the future, but also the likely consequences of man's intervention in the process being forecast. With the big and powerful machines used today, special care must be exercised in forecasts as regards the biosphere — it is to be always borne in mind that errors in the estimates of future trends might lead to implorable consequences some of which have already been mentioned.

It may be asked: what is the time limit for a forecast to remain scientific? When does a forecast degrade to wishful thinking? Or we may ask: is a forecast for any period scientific if an intervention is needed to make it good?

We shall take up these and other questions in connection with the power industry. It is to be stressed that in forecasting adequate account must be taken of social factors, because no power-system expansion can validly be predicted without consideration of how and where industry is to develop and where towns and cities are to be built. Apart from attention to the technical and economic aspects of a power project, be it a station or a transmission line, it is important to bear in mind biospheric, social and demographic matters. Thus, a power-system forecast is always a large and complex problem on a state-wide scale. It may be argued that forecasting has little sense as it handles matters of the future on the basis of the present. For the future is always uncertain! It has been shown more than once, however, that when done with sufficient rigour, forecasts are sufficiently correct. For example, the forecasts which appeared in the press in 1906 materialized 70 to 75%, and those made in 1915-1920 were 80% true. General scientific forecasts (not only in the power industry) have been moderately in error. For example, in the 1950s the Rand Corporation, USA, in its

forecasts based on trend analysis and a poll of a large number of experts predicted for the 1960s the use of a heart-and-lung machine in operations on the heart, soft landing on the Moon, manned space flights, the use of fuel cells on a commercial scale, the advent of vertical-takeoff-and-landing (VTOL) aircraft, the use of antimissile missiles, and gas turbines.

How forecasts relate in terms of the time span covered, and the economic detail given can be shown in per unit terms in a diagram like that of Fig. 1.1

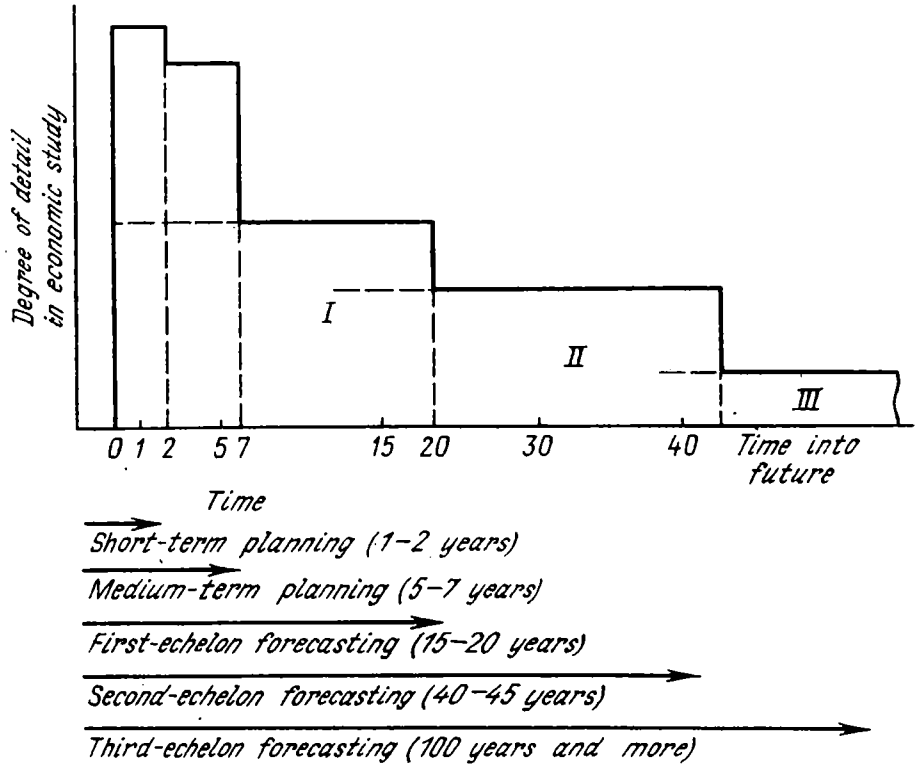

Degree of detail in economic study
I
II
III
0 1 2 5 7 15 20 30 40 Time into future
Time
Short-term planning (1–2 years)
Medium-term planning (5–7 years)
First-echelon forecasting (15–20 years)
Second-echelon forecasting (40–45 years)
Third-echelon forecasting (100 years and more)

Fig. 1.1

There have been wrong forecasts, too. In 1878, William Pierce, chief engineer of the Post Office, UK, turned down Edison's application for a patent covering the electric lamp, arguing that electric light in the homes was an obvious misconception if not a complete absurdity and it would never be used. Professor Pickering in 1910 insisted that flying machines would never cross the ocean, although Bleriot had crossed the English Channel the year before. In his time Heinrich Hertz saw no use for electromagnetic waves. Already in our time, because of a wrong forecast on the use of magnetic discs as a storage medium, the IBM Corporation, USA, decided against making them. Another company in the computer field saw the matter differently and gained a considerable lead on the IBM. It costs much for the IBM to 'stiffle' the competitor and take over the manufacture of discs.

As is seen, wrong forecasts in capitalist society mean heavy losses for the companies concerned to a point where they may be pushed back or even ruined by the wiser rivals. Of course, forecasts dealing with major objectives, both technical and economic, must take proper account of the political climate.

The *second time aspect* is optimization of power-system planning. This involves shorter time spans: 5, 7, 10 or 15 years. Planners deal with more realistic data than forecasters. A five-year plan is based on the development already existing and not on what is only earmarked for realization. Every problem is reduced to its bare necessities so that it can be formulated and solved mathematically, whereas mathematical formalism is not always applicable to forecasts.

In planning, the factors that cannot be incorporated in a mathematical formulation (and such factors are many) are taken care of as constraints or as side conditions not taking part in analysis. An appropriate example can be quoted. Some 25 to 30 years ago, when a plan was put forward to build a hydro station on the river Nurek, USSR, there was a strong body of opinion that the project was premature, and the opponents proved their point by technical and economic calculations. Yet the project was launched because it had demographic, social and political implications. Many more examples could be cited where apart from technical and economic considerations demographic, political, social and defence aspects have had to be reckoned with (for example, the choice of sites for major power stations, water reservoirs, etc.).

How forecasting and planning differ is a question which is not at all simple to answer. Although they have much in common, planning and forecasting problems are formulated mathematically and solved in a different manner. However, the difference tends to diminish because when planning for the use of new apparatus, machines, or materials one has to look into a rather distant future. For example, it took 60 years for the electric motor to move from the inventor's drawing board into use, 50 years for the telephone, 40 years for the tape recorder, 30 years for the zipper, 11 years for nylon, 18 years for X-ray apparatus, 15 years for TV and radar, and as little as 5 years for the nuclear reactor and two years for the solar battery. All of these inventions had had to wait a fairly long time before acceptance by the user (except black-heel stockings and the mini-skirt which, if they may be rated as inventions, won immediate recognition). It is to be noted, however, that this lead time is progressively reducing. In other words, the difference between forecasting and planning is decreasing. That is, the first time aspect, forecasting, and the second time aspect, planning, tend to come closer together for all the difference that exists between them.

The *third time aspect* is short-term planning and designing which covers developments within two or three years.

The *fourth time aspect* is annual or seasonal planning. On a diminishing scale, it is followed by planning by the month and by the week (the *fifth*

time aspect). Finally, there is the *sixth time aspect* with planning by the day, by the hour and, probably, 'instant' planning and, accordingly, optimization of power dispatching.*

Thus, we have defined six time aspects. The number of geographical aspects or sections, may be any, because a section can be cut through a large system at a station, a substation, a line, a regional system, or across the country. Recently, it has been proposed to set up a transcontinental line to connect Asia (Siberia with its huge sources of energy) and Western Europe (France). Such a line would give a sizeable gain in terms of peak power**, because the difference in time between Siberia and France is 8 hours. What is surplus power in Siberia at night would supply peak loads in France. With regard to this line, we may well speak of a transcontinental section. In short, geographical aspects may differ widely.

Forecasting differs in a well-defined way from planning in objectives and methods. On the other hand, all forms of planning, that is the time aspects, have much in common, above all *optimization.*

Optimization. With regard to a power system, optimization seeks to formulate the system state taken as being the best from certain considerations, in a mathematical form known as the objective function and then to minimize (or maximize) this function.

A feature common to all practical applications of optimization is incomplete design data. Thus, one is forced to make decisions under conditions of uncertainty. In planning, the factor of incomplete information is less decisive. For example, daily planning and optimization for a hydro station, a thermal plant or a system as a whole is based on more dependable information although it remains probabilistic and fluid.

In forecasting, information about the projected system is more limited in scope than that about an existing system. Yet, although information about an operating system has a degree of credibility by several orders of magnitude higher than that about a projected system, its credibility is not absolute, because input data are of necessity deformed in the course of acquisition and reduction. In fact, we have to be content with probabilistic data even when we deal with a running system. These data can be regarded as deterministic only arbitrarily. For example, the performance of steam boilers will degrade with time because, say, the thermal conductivity of the boiler tubes is reduced (by scale formation etc.). In other words, what we actually have are probabilistically varying data, but we often take them as being deterministic.

There is a marked difference between planning and forecasting in the way the acquired data are processed and in the mathematical techniques involved, as these are different for each of the six time aspects.

The objectives in the first time aspect, that is, forecasting, are to stake

* No accepted terminology exists, unfortunately. Sometimes, forecasting is applied to the calculation of operating parameters for the next day.

** We leave out the economy of such a power transmission scheme.

out, at least tentatively, the road or roads to be taken, to select valid quantitative indicators, and to assess how much capital should be put in a particular industry or project. For example, the alternatives are to put money in MHD or super-big generators. If we choose the latter, the next question may be how big they should be. As a tentative estimate, we may assume that by the year 2000 we must have generators rated at 6000 MW apiece. This immediately leads to the question as to how valid this estimate is. If it is valid, we must work so as to have such generators by the year 2000. This involves investment in research and development work on these generators and denial of investment to some other, competitive projects. Similar reasoning may be applied to power transmission lines. Where ought we to put in money? Should we set sights on transmission lines using compressed gas in conduits, or cryogenic lines, or d.c. lines? In matters of where to put in money and how much it is vitally important to avoid false estimates of future capabilities. This is the prime task of forecasting.

In the second and third time aspects, the tasks are to plan capital investments for specific projects, after the general or main directions of development have already been mapped and we have to decide what to build and when. We also face the problem of priorities among the projects at hand — which is to be given money, completed and put in operation first, and how its commissioning dates are to be correlated with those of the other projects.

Planners usually feel confident when they handle economic and technical data, although the incompleteness and probabilistic nature of information complicate the matters, too.

In the fourth and sixth time aspects, the top objective is to run the system in such a manner as to supply power of the right quality and in the right quantity at minimum cost. The task may be formulated as follows: given, a running power system; required, to control it so as to supply the consumer with the specified quantity of energy of the desired quality. Also, energy must be generated in a manner which would cut down the cost for the system as a whole. The generating set must be loaded so as to keep losses in the plant and in the lines to a minimum. Again, mathematically the problem reduces to minimizing the objective function describing the cost of the power delivered to consumers, or to maximize the objective function related to power generation. In the final analysis, we seek to find conditions under which a minimum call would have to be made on capital, fuel and manpower resources. However, certain constraints have then to be imposed on the mathematical models, because, apart from economic considerations, the objective function would not usually contain terms for all kinds of limitations. For example, the cost of electricity in the United States today is half as much again as it was until recently, and the consumer has to pay a 'scarcity' price, but not because the energy resources have become scarce. The point is that some states have passed laws prohibiting the use of fuel causing excessive air pollution (by emitting too much ash, sulphur, etc.). This has led to an increased demand for high-

quality fuels and, as a corollary, to higher prices. Fuel suppliers have raised the price on fuel, and power utilities have done so on electricity. This example illustrates the effects of constraints caused by legislation and by the desire to make a bigger profit, a feature most characteristic of capitalist society.

In short, problems of power-supply optimization are subject to an ever increasing number of constraints. These constraints complicate mathematical solution and cause the gap to close between the tasks of forecasting, planning and short-term system state optimization which fall in the domain of system control.

In the Soviet Union, matters of management and forecasting are at the focus of public attention, and they were dealt with at length by the 24th Party Congress. To quote L. I. Brezhnev, streamlining of management is not a single action, but a dynamic process of solving the problems posed by life. It is to be recalled that the Directives for the ninth five-year plan devoted a whole section to streamlined management and control, and that the country's leading agencies concerned with planning and scientific management are engaged in the forecasting of the availability and consumption of resources (Fig. 1.2)

Management of the power industry may be defined as a purposeful action on a large man-made system, which the power industry is, so as to achieve the desired goal. This management is effected in all the time aspects and must rely on cybernetic methods so that the purposeful action taken in the circumstances dictated by the environment can lead to the sought goal in an optimal manner. In this sense, the optimization of management and control, implies the optimization of forecasting, planning and expansion*.

Speaking of optimization, it should be noted that in a socialist economy the most favourable alternative plan is that which promotes the functioning and development of socialism as a social system, that is, which helps with the solution of problems facing society in the pursuit of a specific goal — that of building communism. In fact, the principal criterion of optimality in the USSR is how much a given alternative advances this final goal. It should be admitted, though, that this indicator is too general for practical, day-to-day needs. Apart from this main indicator which is at the base of socialist evolution, more concrete and specific indicators are needed, directly applicable to current forecasts, plans and, the more so, to current design work and operation.

Although the capitalist world is governed by the laws of supply and demand (the market economy), there are examples of the efficient management of individual factories, concerns and even international corporations. In some cases, several countries pool their resources, as Canada, Italy and, partly, France have done in the power industry, in order to tackle problems

* Optimization has its origin in the Latin 'optimus', meaning most favourable (e.g. conditions for growth, reproduction, etc.). When used without reference to a specific thing, it is meaningless, although some use it as a substitute for 'good', which is wrong.

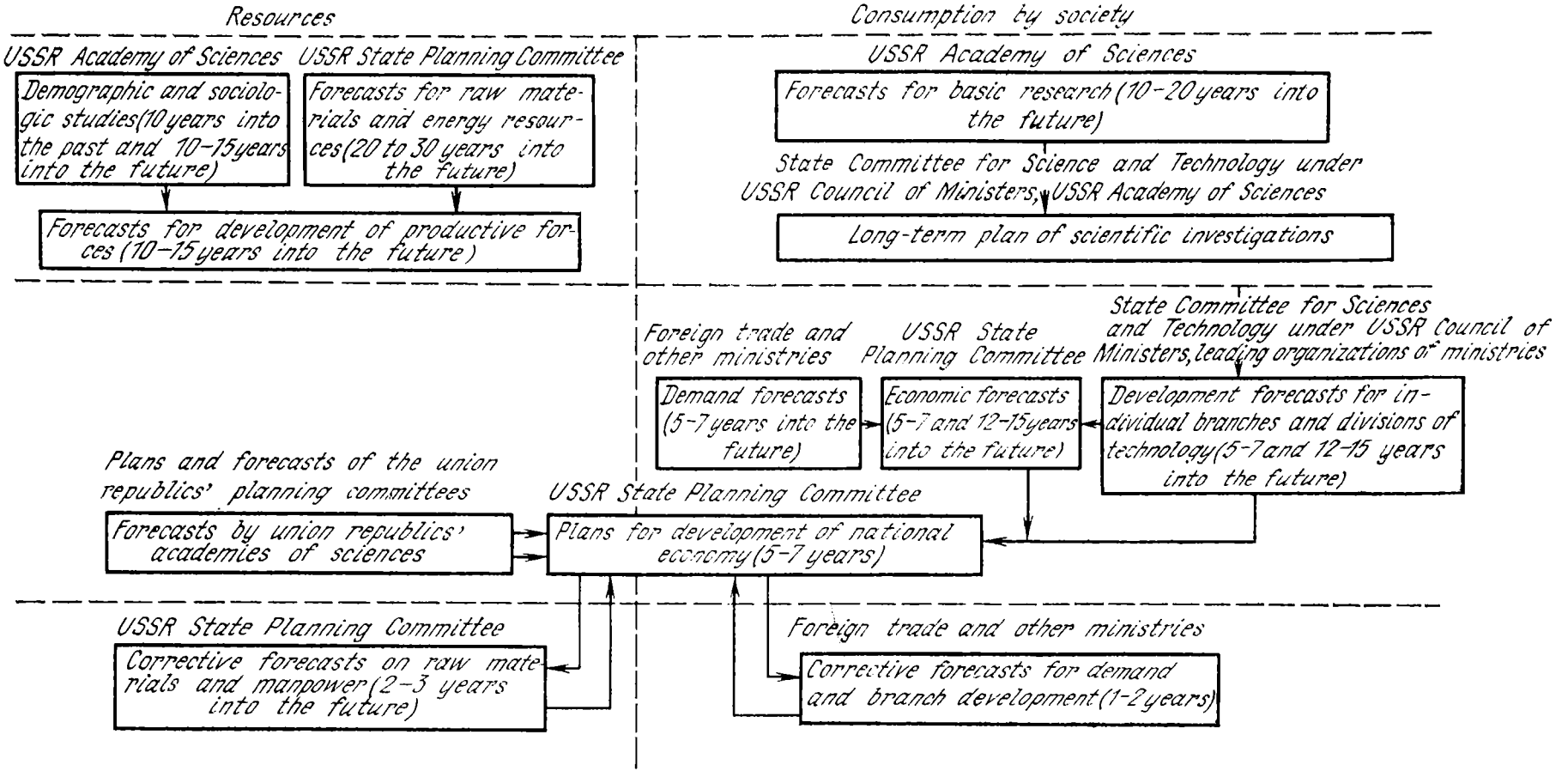

Resources
Consumption by society
USSR Academy of Sciences
USSR State Planning Committee
Demographic and sociologic studies (10 years into the past and 10–15 years into the future)
Forecasts for raw materials and energy resources (20 to 30 years into the future)
Forecasts for development of productive forces (10–15 years into the future)
USSR Academy of Sciences
Forecasts for basic research (10–20 years into the future)
State Committee for Science and Technology under USSR Council of Ministers, USSR Academy of Sciences
Long-term plan of scientific investigations
Foreign trade and other ministries
USSR State Planning Committee
State Committee for Sciences and Technology under USSR Council of Ministers, leading organizations of ministries
Demand forecasts (5–7 years into the future)
Economic forecasts (5–7 and 12–15 years into the future)
Development forecasts for individual branches and divisions of technology (5–7 and 12–15 years into the future)
Plans and forecasts of the union republics' planning committees
USSR State Planning Committee
Forecasts by union republics' academies of sciences
Plans for development of national economy (5–7 years)
USSR State Planning Committee
Foreign trade and other ministries
Corrective forecasts on raw materials and manpower (2–3 years into the future)
Corrective forecasts for demand and branch development (1–2 years)

Fig. 1.2

associated with operation at super-high voltages. In this particular example, Italy tested transmission line supports which were then set up in Canada. This was management on an international scale. Some of the capitalist system's capabilities and experience in industrial management are worth close study and attention, although a truly and fully scientific management of society and public economy is incompatible with the economic basis of capitalism where the private ownership of means of production cannot but lead to anarchy in a country's economy as a whole.

In this respect, the Soviet Union has enormous potentialities, although there are also some limitations. Some of these limitations stem from the fact that the effectiveness indicators used for specific calculations are too narrow and cannot therefore yield correct results always, whereas the overall system indicator is too general for specific needs. For example, the usual quantitative economic indicators cannot give a correct measure of the effectiveness of defence efforts; very often we run into a difficulty when we attempt to assess the social and political payoff from purely technical forecasts. We, in the USSR, have to reckon with a multitude of factors or constraints apart from those related to the biosphere, etc. The power-system engineer has of late been running into these matters on an ever increasing scale.

In passing to the specific problems of forecasting, a few remarks are in order, concerning the general definitions given earlier. The power system as a whole may be pictured as shown in Fig. 1.3, where the bold solid lines

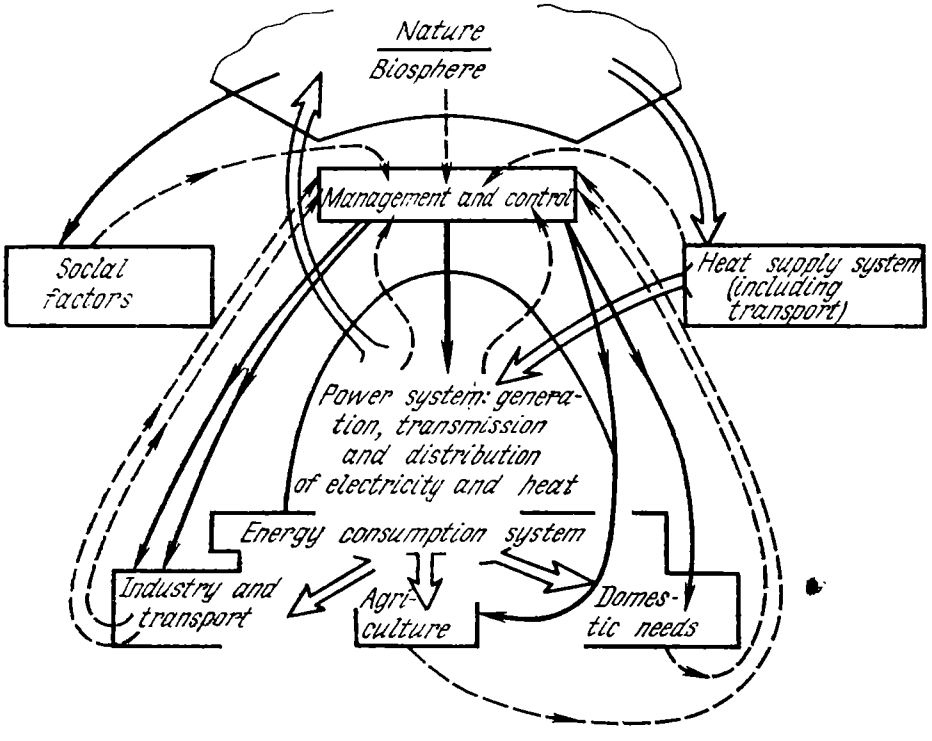


Fig. 1.3

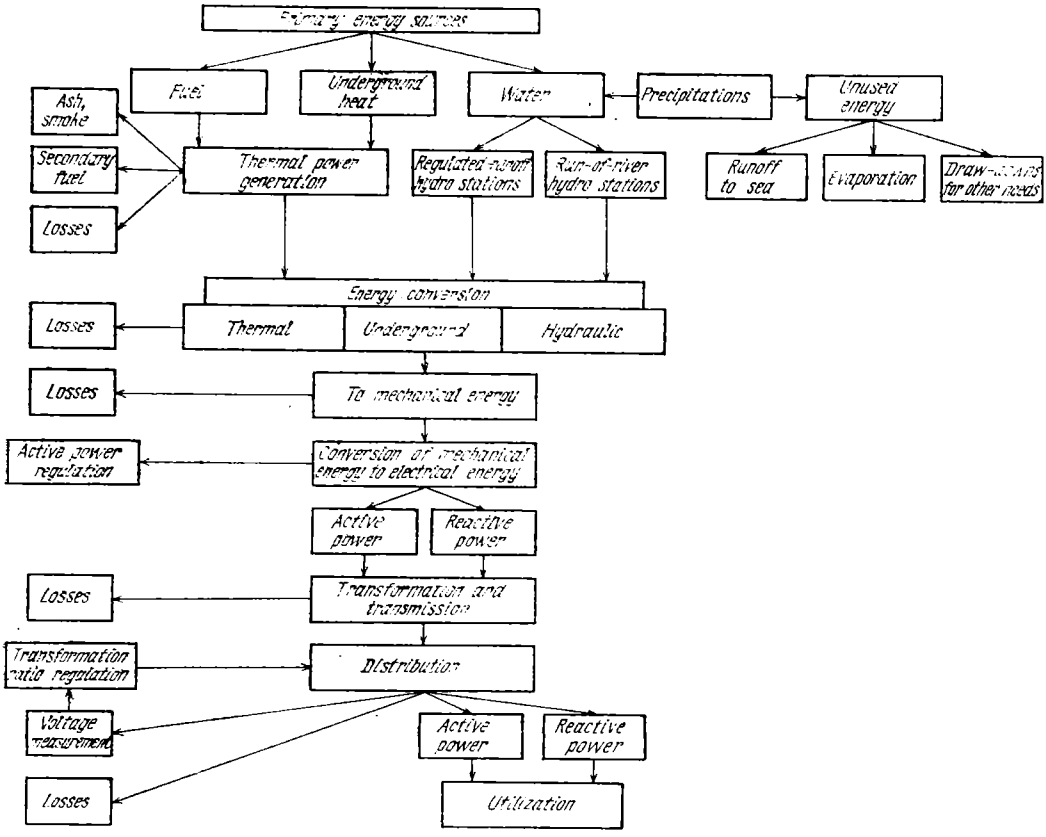


Fig. 1.4

mark power and fuel flows and connections, the fine solid lines represent control and management, and the dashed lines stand for information paths. In an analysis of the power system proper, the many external systems which are coupled to it to form, between them, a pan-energy system, will be treated as the surroundings. The structure of power generation and utilization may be depicted as shown in Fig. 1.4, and the geographical breakdown of the power industry as shown in Fig. 1.5. Speaking of cybernetic control of system expansion, we shall consider not simply a system which is expected to supply a specified quantity of energy, but a system which should do so subject to the constraints imposed by the surroundings. This is a cybernetic approach.

The system may be decomposed into a series of levels. One such decomposition of the system hierarchy made up of a level representing the country's pan-energy system (which is to be managed optimally in the wider sense defined earlier), a regional level, a load-centre level, and an undertaking level, is shown in Fig. 1.6*. The solid lines represent vertical subordination and the dashed lines depict interactions and interoperation.

* After Academician L. A. Melentyev.

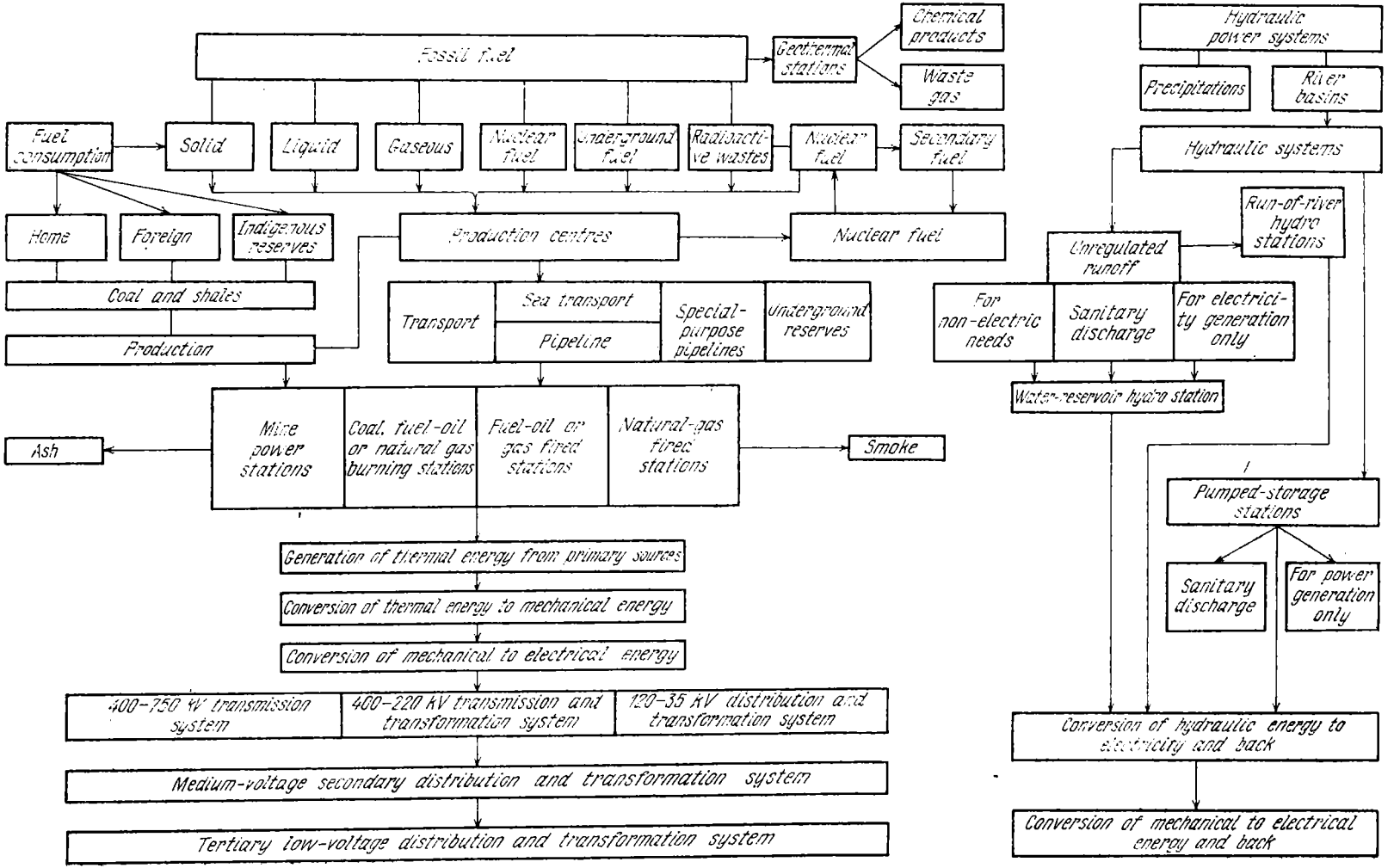

Fossil fuel
Geothermal stations
Chemical products
Waste gas
Fuel consumption
Solid
Liquid
Gaseous
Nuclear fuel
Underground fuel
Radioactive wastes
Nuclear fuel
Secondary fuel
Home
Foreign
Indigenous reserves
Production centres
Nuclear fuel
Coal and shales
Production
Transport
Sea transport
Pipeline
Special-purpose pipelines
Underground reserves
Ash
Mine power stations
Coal, fuel-oil or natural gas burning stations
Fuel-oil or gas fired stations
Natural-gas fired stations
Smoke
Generation of thermal energy from primary sources
Conversion of thermal energy to mechanical energy
Conversion of mechanical to electrical energy
400–750 kV transmission system
400–220 kV transmission and transformation system
120–35 kV distribution and transformation system
Medium-voltage secondary distribution and transformation system
Tertiary low-voltage distribution and transformation system
Hydraulic power systems
Precipitations
River basins
Hydraulic systems
Run-of-river hydro stations
Unregulated runoff
For non-electric needs
Sanitary discharge
For electricity generation only
Water-reservoir hydro station
Pumped-storage stations
Sanitary discharge
For power generation only
Conversion of hydraulic energy to electricity and back
Conversion of mechanical to electrical energy and back

Fig. 1,5

The country's pan-energy system (Fig. 1.6) incorporates an electric power system, an oil supply system, a gas-supply system, and a coal-supply system. They make the four sub-systems of the overall system. No one can be examined in isolation from the remaining three and from the overall system, because of the external flows and connections shown in Fig. 1.3. At the regional level, which comes next, there is, again, a pan-energy system of its own, but the external connections act differently. The next lower levels are those associated with the load-centres and individual undertakings; these may be power stations, boiler plants, oil fields, gas lines, collieries, etc. At the base of this hierarchy are consumers of electricity, steam and hot water, oil and oil products, gas and coal.

To sum up, we have to deal with matters at four levels where energy is

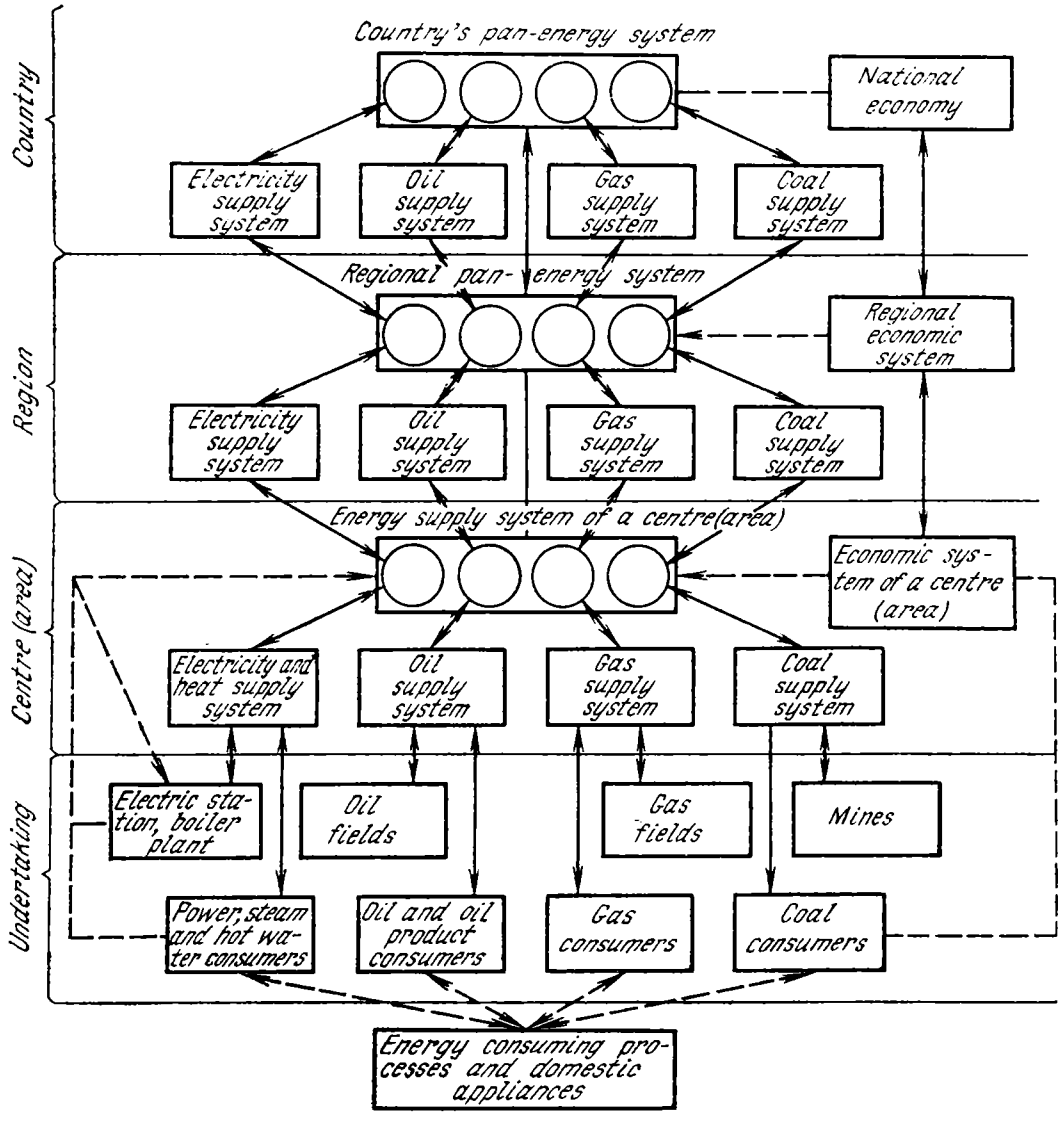


Fig. 1.6

produced and utilized. At each of these levels, forecasts, plans and operational control should be managed so as to strike a balance, or rather, to find a compromise, between the various indicators, such as capital investment and operating costs. These indicators should be set up in mathematical equations whose solution will lead to the optimization of the system-wide indicators, although some of the subsystems may have to be run at a sub-optimal level.

Usually, optimization is conceived as the minimization or maximization of a variable. This is an over-simplification. In fact, there is a multi-extremal function of several variables, so that the search for a point defining the minimum of a single variable is an abstraction. Practically, we may speak of an optimization domain appearing as a surface in a simple case, and of the manner in which this domain can be approached, subject to a multitude of constraints. These constraints may be viewed as if dissecting the optimization domain into sub-domains which cannot be defined or isolated by a rigorous mathematical formulation. These sub-domains can be identified by the additional factors that are decisive in a given set of conditions. Yet, the solution converging to an optimum point is useful, especially at the early stages of management and control. Above all, the system saves fuel, although the calculations are imperfect, as they ignore some important factors. Obviously, such calculations serve a useful purpose despite the fact that they are approximate. It may be added, though, that little can be gained by making them more precise at a later time.

1.2. Forecasting Methods

We shall take up specific tasks of forecasting, while leaving matters of planning and control, that is optimization, of an existing system for a later chapter.

Forecasts may be different: for science, for technology, and for industry. For all the difference, however, they cannot be fully separated, for the gap between science and production is closing. Very often, scientific problems are solved by carrying out studies on an industrial scale.

We find a brilliant example of a social and technical forecast in the GOELRO plan: it provided for development in both industry and agriculture on the basis of advances and expansion in the power industry and large-scale electrification.

Today, forecasting in the power industry is particularly important. In the nuclear power industry, for example, we need sound forecasts on the disposal of nuclear wastes, the production and utilization of nuclear fuel. In assessing the impact of power generation on the biosphere, it is to be remembered that the thermal plants of the year 2000 would need huge amounts of cooling water unless entirely new designs of cooling towers are engineered. If we assume that generating unit capacity at thermal plants remains as it was in 1965-1970, the thermal plants would take up an area the size of Belgium. Accordingly, forecasts are needed on water usage,

cooling-tower designs, turbo-generator designs, and the materials and components that would increase generating unit capacity and enhance the reliability of generating plant. Unless radical changes are made in thermal-plant configuration, cooling water would carry away with it about as much heat as the whole of the country is estimated to utilize in the year 2000. This heat, if released, would be able to change the climate over large areas, and this change needs a forecast of its own. The monstrous figures quoted in existing forecasts on water and air pollution make it imperative to look for entirely novel designs of power generating installations. Thus, it is not enough simply to forecast power plant capacities; much thought must also be given to how to build them in a new way and how to run them in the changed conditions. Taking account of all interrelationships and approaching all matters cybernetically, we must look for revolutionary technical decisions and not simply forecast the quantitative expansion of the power industry. In particular, it is important to increase capacity per generating unit. Already now, plans are under way to build power generating units rated at 3000 MW apiece, that is, each equal to three units of the Kuibyshev Hydro. At the same time, it is specified that each unit be of moderate size in order to save land. The difficulties posed by the discharge of hot cooling water make it imperative to forecast the likehood of using improved closed-circuit cooling-water schemes as these would not release too much heat into the biosphere. Stressing it again, forecasts should not only state how many generating units should be built to generate so much power, but also outline what should be done to advance science and technology. Obviously, this can only be done by striking a balance between many and widely differing and conflicting factors — economical, technological and scientific.

No planning is possible without forecasting. Yet, many instances might be quoted where no reliable forecasts had been made prior to planning and where the returns were therefore poor. Optimum decisions can only be made through a comparison of a reasonable number of alternative forecasts and expansion plans. Yet, again, cases might be quoted where a design or an expansion plan would be compared with only one competing alternative known in advance to be inferior, with the sole purpose of mock optimization.

To make forecasts reliable, consistent input data must be used for costing, accounting, analysis, control, etc.

Typically, current forecasts set as their target time the year 2000. This is true of practically all forms of forecasting in both the Soviet Union and other countries.

Somewhat arbitrarily, all forecasting methods may be classed into three groups: *extrapolation*, *expert opinion* and *modelling*. Between them, the three groups total close on a hundred distinct methods. It would be hardly practicable to dwell on each separately, the more so that there is no rigid delineation between the techniques used*, and often one method liberally

* A more detailed classification is given in Fig. 1.7.

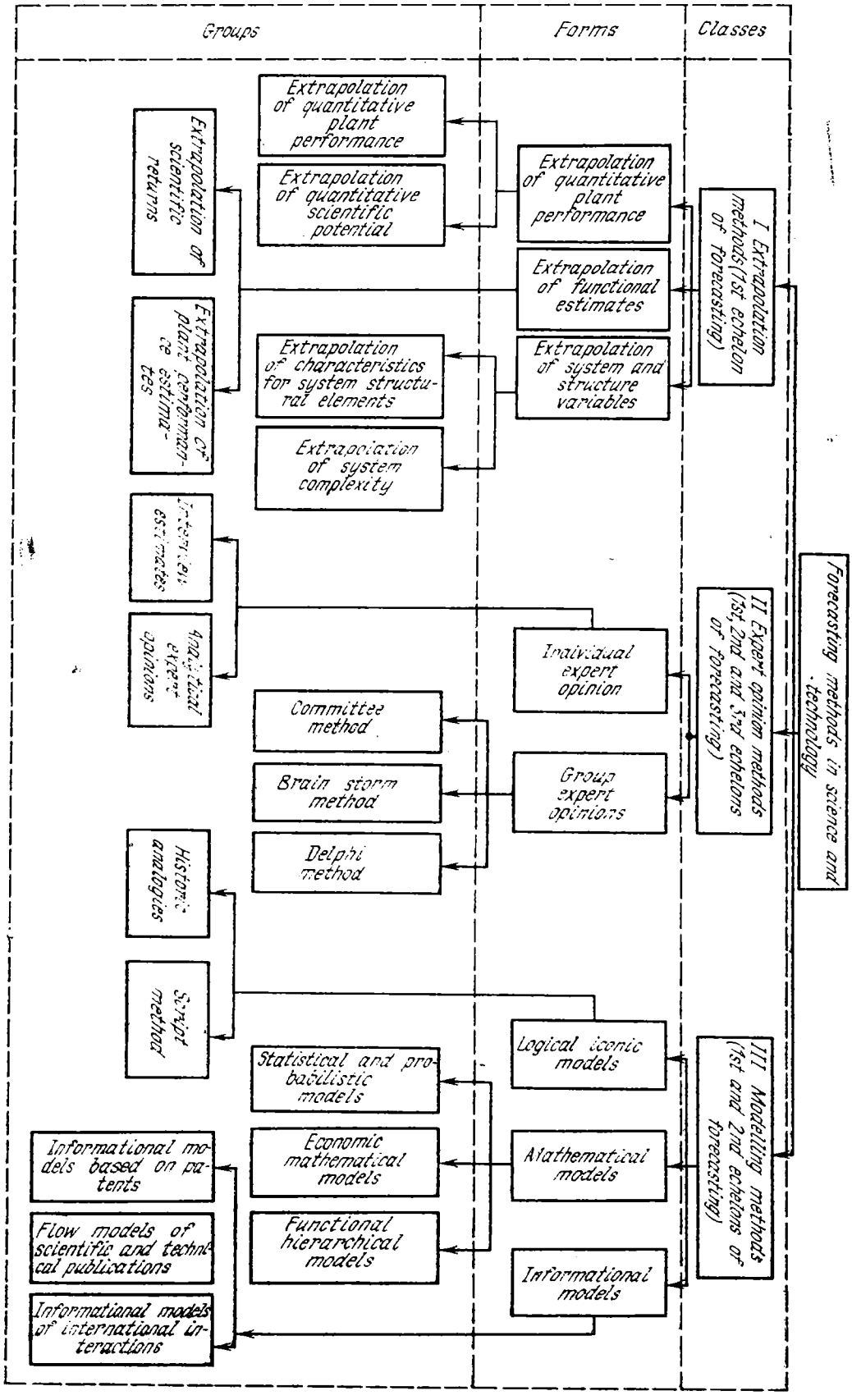

Groups
Forms
Classes
Extrapolation of quantitative plant performance
Extrapolation of quantitative scientific potential
Extrapolation of scientific returns
Extrapolation of quantitative plant performance
Extrapolation of functional estimates
I Extrapolation methods (1st echelon of forecasting)
Extrapolation of plant competitive estimates
Extrapolation of characteristics for system structural elements
Extrapolation of system complexity
Extrapolation of system and structure variables
Forecasting methods in science and technology
Individual expert estimates
Analytical expert opinions
Individual expert opinion
II Expert opinion methods (1st, 2nd and 3rd echelons of forecasting)
Committee method
Brain storm method
Delphi method
Group expert opinions
Historic analogies
Script method
Logical iconic models
Statistical and probabilistic models
Economic mathematical models
Functional hierarchical models
Mathematical models
Informational models
III Modelling methods (1st and 2nd echelons of forecasting)
Informational models based on patents
Flow models of scientific and technical publications
Informational models of international interactions

Fig. 1.7

borrows some techniques from another (Fig. 1.7). Our discussion will therefore be limited to the most typical methods.

The steps involved in extrapolation are:

(1) detection of a trend in the parameters;

(2) evaluation of the functional relationships;

(3) estimation of system and structure parameters.

The expert opinion methods can be further subdivided into two subgroups as follows:

(1) individual expert methods;

(2) expert group methods which also include the Delphi methods (so called by analogy with the town of Delphi in ancient Greece where Apollo used to issue his oracles by the mouth of his priestess the Pythia).

As regards modelling, there may be logical, mathematical and information models.

Extrapolation. With extrapolation, the future is treated as an extension of the past. A forecast is made by projecting into the future the events and states that occurred over a time period Δt, that is, their trend in the past into the future (Fig. 1.8). Several modifications of the extrapolation method exist, such as based on parametric estimation, trend curves, and the special properties of the polynomials used in extrapolation.

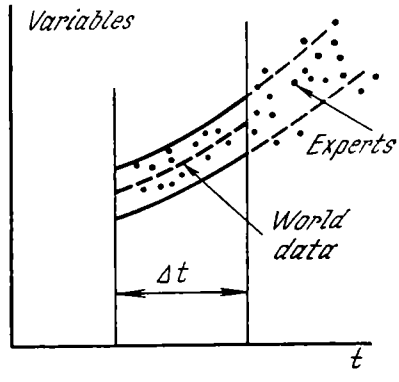


Fig. 1.8

Of late, extrapolation has been severely criticized by those who question the validity of projecting past events into the future. Yet, there is a belief that extrapolation can produce forecasts for 12 to 15 years ahead with an error of $\pm 15\%$. Quite a number of forecasts for 10 to 15 years in advance have worked well. An example is the forecast on the speed, capacity and cost of computers for the year 1975 made on the basis of data observed over a time period starting in 1955 and on the assumption that the observed trend would remain the same. The follow-up check showed the forecasts to be basically correct.

It must be admitted, however, that extrapolation has been justly criticized for lack of rigour; to improve its efficiency, deeper insight is needed into the logic of a development or the physics of an event. Otherwise, a forecaster may well end up with preposterous results. According to such unsound extrapolation, the number of scientists in the year 2000 might exceed the total world population, there might be over a million artificial satellites orbiting the Earth in 1990, and the vehicles of the year 1990 might travel faster than light. Obviously, simple extrapolation of a short trend onto a long time period in the future is unreliable. No statistical technique can bring out internal relationships or probe deep into the physics of events. This is why additional analysis is essential.

As an example, Figure 1.9 illustrates a forecast for the practical use of the electromagnetic spectrum. As is seen, practical use will be made of gamma-rays in the year 2000, and of cosmic rays in the succeeding years.

With extrapolation, time is laid off as, say, ordinate, and the developments already existing are plotted as dots along the other (say, x) axis. Next, extrapolation proper is carried out as follows:

(1) a hypothesis is advanced on the mechanism by which the development may proceed;

(2) indicators are chosen that are best related to the development being forecast, and units of measurement are assigned to them;

(3) data are collected and grouped on the indicators chosen;

(4) these data are reduced statistically and a trend line (or lines) is determined. This is usually done by plotting the observed data along the log axis of a semi-log paper (with time plotted along the other axis uniformly). This yields a field of dots. Then using the least squares technique, a curve is constructed, defined as

$$\ln X = a_0 + a_1 t + a_2 t^2$$

This form of the trend curve is chosen because it usually gives a good fit to the points representing the observed data and is simple to determine;

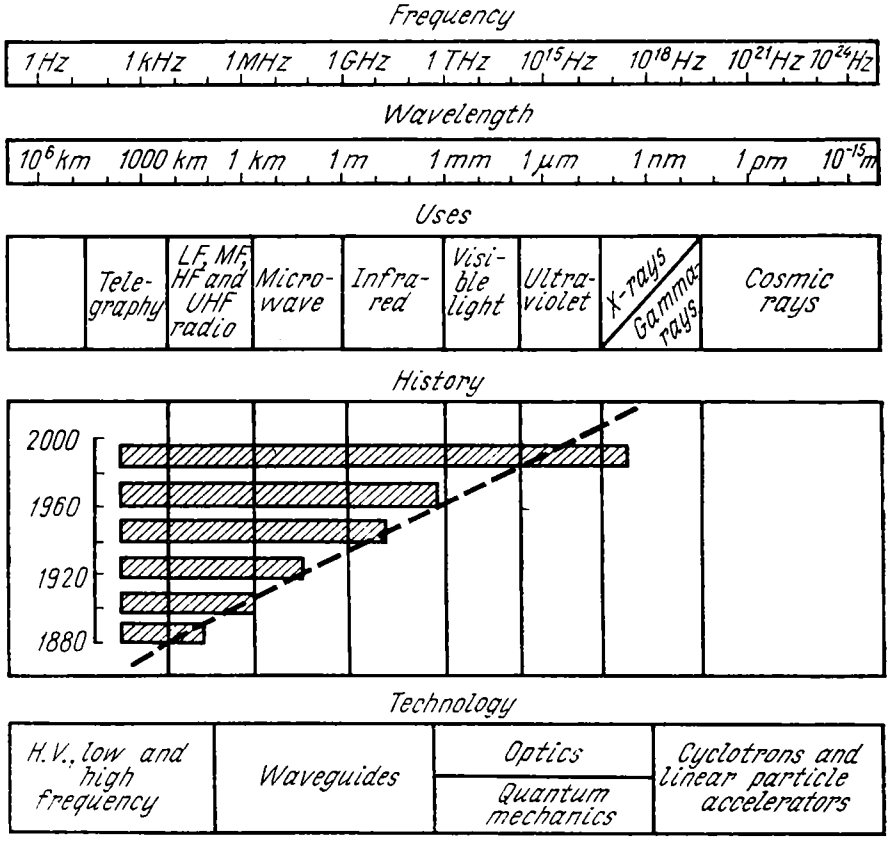


Fig. 1.9

(5) the necessary calculations are made;

(6) a final decision is made on the limits of extrapolation. The need for such limits has been proved by the examples quoted earlier (see p. 65);

(7) the results are reviewed in the light of the mechanism hypothesized at the outset, and the necessary revisions and refinements are made so that when work is completed there is a good measure of agreement between step (1) and steps (5) through (7).

Truly scientific forecasting has nothing to do with blind faith in statistics. Even if a forecast yields an absurdity, this is an indication that some change is needed in the incipient trend.

In many forecasting methods, a major problem is the choice of a minimal number of indicators needed to make credible forecasts. This is where some help may come from *correlation analysis*, by which the forecaster can retain within his field of view only the leading indicators closely related to the problem and discard all the minor characteristics. In mathematical statistics, a measure of this closeness is given by a *correlation coefficient*. There may be the pair correlation coefficient, r_{yx}, the partial correlation coefficient, r_{xk}, and the multiple correlation coefficient, R.

The coefficient of pair correlation shows how closely the leading indicator being forecast, y, depends on any one of p independent variables $x_1, x_2, \ldots, x_p$ (say, x_1). The coefficient of partial (or net) correlation gives a measure of such closeness between any pair of independent variables, say $x_1=f(x_2)$, $x_2=f(x_3)$, $x_3=f(x_4)$, etc. The coefficient of multiple correlation gives a measure of closeness between the leading indicator being forecast and a combination of independent variables including $x_1, x_2, x_3, \ldots, x_{p-1}$, up to x_p.

The equations describing the various coefficients of correlation can be found in books on mathematical statistics. We shall only give the equations for the coefficients of pair correlation, used in forecasting to determine the closeness of relationship between a leading indicator and each of the factors included in the set:

$$r_{yx}=\frac{\sum_i (y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum_i (y_i-\bar{y})^2 \sum (x_i-\bar{x})^2}}$$

where $\bar{y}$, $\bar{x}$, y_i and x_i are, respectively, the arithmetic means and specific data related to the leading indicator being forecast and to one of the factors, or independent variables, tested in turn.

When the relationship between y and $x_1, x_2, x_3, \ldots, x_p$ is nonlinear, a measure of relationship is given by the correlation ratio

$$\eta=\sqrt{\frac{\sum_{j=1}^{m} v_i(y_j-\bar{y})^2}{\sum_{i=1}^{n} (y_i-\bar{y})^2}}$$

where m = number of grouping intervals
v_j = number of points falling within the jth grouping interval
$y_i = \sum_{i=1}^{v_j} y_{ij}/v_i$ = average of the jth observation interval
n = number of observations (experimental points)
$\bar{y} = \sum \bar{y}_j v_j/n$ = total average value of the leading indicator.

For a linear relationship, the correlation ratio is identically equal to the coefficient of correlation, r_{yx}. When the relationship is nonlinear, the correlation ratio is defined as

$$\eta^2 > r_{yx}^2$$

Generally, there is a closeness between the leading indicator being forecast and an appropriate independent variable if the coefficient of correlation or the correlation ratio approaches $+1$ or -1. The '+' sign of the coefficient of correlation applies when the leading indicator is directly proportional to the independent variable, while the '$-$' sign is an indication of inverse proportionality.

The dependence of y on $x_1, x_2, \ldots, x_p$ is nonlinear if the coefficient of correlation and the correlation ratio are less than unity. However, if they are zero, this ought not to be taken as an indication that y does not depend on $x_1, \ldots, x_p$ or that the relationship is nonlinear. Whether this is so should be verified through an additional analysis.

The values of r_{xy} and η markedly depend on the amount of data (or the size of the sample) available to the forecaster. In this connection, it is important to establish how representative a given sample is. Practically, this is done by testing the coefficient of correlation with the theoretical distribution, t_ν (Student's distribution). In books on the theory of probability, Student's distribution is defined as a function of the sample standard deviation (for levels of significance $\alpha = 0.1, 0.05, 0.025, 0.01, 0.005$ and 0.001) and the number of degrees of freedom, $\nu = n-p-1$ (where n is the number of observations and p is the number of independent variables). Accordingly, we may write:

for large samples ($n > 25$)

$$r\sqrt{n-1}/(1+r^2) > t_\nu$$

for small samples ($n < 25$)

$$r\sqrt{n-2}/\sqrt{1-r^2} > t_\nu$$

In general form, the dependence of the leading indicator being forecast, y, on the independent variables $x_1, x_2, \ldots, x_p$ may be defined by a regression equation of the form

$$y = a_0 \sum_{i=1}^{p} (a_{1i}x_i + a_2 x_i^2 + \ldots + a_{ni}x_i^n) + a_m t$$

where a_0 = constant term of the equation

$a_1, a_2, \ldots, a_p$ = coefficients of regression of p variables $x_1, x_2, \ldots, x_p$
t = time variable
a_m = coefficient of the time variable

If, in the regression equation, all terms are of power one, we have *linear regression.* If, on the other hand, at least one of the terms is of a greater power, this is a case of *nonlinear regression.*

When the leading indicator is a function of only one independent variable, the forecaster analyses a *single-variable model;* otherwise he has to deal with a *multi-variable model.*

In linear multi-variable form, the regression equation of a statistical problem may be written as

$$y-\bar{y}=a_1(x_1-\bar{x}_1)+a_2(x_2-\bar{x}_2)+\ldots+a_p(x_p-\bar{x}_p)$$

or

$$y-\bar{y}=\sum_{i=1}^{p} a_i(x_i-\bar{x}_i)$$

where $\bar{y}$ and $\bar{x}$ are the arithmetic means of the leading indicator being forecast and of the independent variables, y and x_i are the specific values (observations) of the leading indicator and appropriate independent variables, and p is the number of independent variables.

The above equations can be derived from the general equation of regression, if all terms are expressed in terms of the constant term

$$a_0=\bar{y}-a_1\bar{x}_1-a_2\bar{x}_2-\ldots-a_p\bar{x}_p$$

and appropriate substitutions are made in the general equation. From the transformed equation it follows that an increment in the linear function, $\Delta y=y-\bar{y}$, is proportional to the increment in the independent variable, $\Delta x=x-\bar{x}$, and is independent of the initial value of the constant term, a_0. This is why the coefficients of linear (or straight-line) regression are called *coefficients of proportionality.* They show how much the mean value of the leading indicator is caused to change as the mean of a given independent variable is incremented by unity while all other independent variables are held constant. In other words, the regression coefficients a_i are *constant* for all values of a given independent variable, that is

$$a_i=\Delta y/\Delta x=\text{constant}$$

Hence, the leading indicator is directly proportional to an appropriate independent variable when $a_i>0$, and inversely proportional when $a_i<0$.

The simplest equation of single-variable linear regression has the form

$$y=a_0+a_1x$$

In this case, the regression line (that is, the line representing the relationship between the two quantities) is a straight line, with a_0 representing the distance from the origin to the point of intersection of the regression line

and the y-axis, and a_1 giving the slope of the regression line, that is, the tangent of the angle which the line makes with the x-axis, $\tan \alpha = a_1$.

A problem may be taken as solved if all coefficients in the general equation (or the coefficients a_0 and a_1 in the equation of single-variable linear regression) have been evaluated.

Of the infinite multitude of regression lines (empirical regressions), the one which fits the observed data best will be that which satisfies the least squares equation

$$\sum_i (y_i - \bar{y}_i)^2 = \min$$

or in words, that line for which the sum of the squares of the deviations of the observed values (the separate points), y_i, of the leading indicator about its central or estimated value, y, as represented by the line of regression (i.e., the squares of the residuals) is a minimum.

By the least squares method, the two coefficients, a_0 and a_1, of single-variable linear regression can be found by solving two equations

$$\sum_i y_i = n a_0 + a_1 \sum_i x_i$$

$$\sum_i y_i x_i = a_0 \sum_i x_i + a_1 \sum_i x_i^2$$

where n is the number of observations (the size of the sample).

Two-variable linear regressions involve the solution of three simultaneous equations to find three coefficients of regression, a_0, a_1 and a_2. Hence, we may conclude that the number of equations needed to evaluate the coefficients of a linear regression is

$$m = n + 1$$

where m is the number of equations and n is the rank of the regression (it may be single-variable, two-variable, etc.).

Very often, forecasters resort to curvilinear regression, especially when they have to deal with single- and two-variable models. Whether or not the use of curvilinear regression is indicated is determined by the value of V^2:

$$V^2 = \frac{(n - m_0) \sum_{i=1}^{m_0} v_i (\bar{y}_i - y_i)^2}{(m_0 - p) \sum_{i=1}^{m_0} \sum_{j=1}^{v_i} (y_{ij} - \bar{y})^2}$$

where n = amount of available data (total number of observations or design data)

m_0 = number of data grouping intervals

v_i = number of data sources (points) falling within the ith grouping interval

p = number of independent variables

$\bar{y}_i$ and y_i = mean and calculated values of the leading indicator being forecast

The value of V^2 thus found is then tested with Fisher's distribution, $F(\alpha, m_0-p, n-m_0)$, which is a function of the probability of deviation (at levels of significance $\alpha=0.1$, 0.05, etc.), the number of degrees of freedom $n-m_0$ (in the numerator) and the number of degrees of freedom m_0-p (in the denominator). Curvilinear regression is indicated for cases where the values of V^2 exceed the tabulated values of $F(\alpha, m_0-p, n-m_0)$, that is, when

$$V^2>F(\alpha, m_0-p, n-m_0)$$

Conversely, linear models should be preferred when

$$V^2<F(\alpha, m_0-p, n-m_0)$$

As a rule, single-variable curvilinear regressions are described by simple functions, notably:

the exponential function

$$y=\exp(a_0+a_1x)$$

the logarithmic function

$$y=a_0+a_1\ln x$$

the power function

$$y=a_0x^{a_1}$$

the parabolic function (a single-variable parabola of order two)

$$y=a_0+a_1x+a_2x^2$$

As before, a problem in regression analysis may be taken as solved if the coefficients (constants) $a_0, a_1, \ldots, a_n$, in the regression equations have been evaluated. Their evaluation is based on a technique known as *linearization*, by which a curvilinear regression is transformed into a linear one.

Linearization is usually performed by techniques known from the statistical analysis and reduction of experimental data*. Most often, this is done by taking logarithms of the functions involved. It is readily seen that by taking their logarithms all elementary approximating functions are reduced to linear form. Once this is done, the regression coefficients can conveniently be secured by the least squares method.

Expert-opinion method. Forecasts by this method involve querying and re-querying leading experts in a particular field. Although by a jocular definition, an expert is a specialist who has made a lot of mistakes, he must actually be a man of profound intuition, extensive erudition, and experience.

Expert opinions may be advanced by *individuals* or by *groups*. In the latter case, experts may be invited to sit together in a session or a questionnaire may be circulated among them.

Of all varieties of individual expert techniques, the *morphological* technique is used most. In effect, this is an exposition of an ordered view on

* See, for example, *Treatment of Experimental Data* by A. Worthing and J. Geffner.

things. The technique seeks to present a systematic review and analysis of the problem at hand and to take account of all known possibilities and the properties of the object by decomposing the problem.

Unfortunately, all individual methods are liable to subjective misconceptions and mistakes. This disadvantage is largely avoided in forecasts based on group opinions.

Among the group opinion methods, resort is most often made to panels of experts. Such panels can work in a variety of ways, each known by a specific name. The most commonly used forms are the brain storm technique, the cross-examination technique, and the court-of-justice technique.

In a *brain-storming session* a group of experts is called together and invited to discuss a problem of interest. Each is encouraged to give his opinion, however 'wild' it may sound — no critical analysis of ideas is permitted, although it is hoped that each 'wild' idea may lead to another and, perhaps, to a radical idea which has merit. A brain-storming session usually lasts for an hour and a half or so, with every opinion recorded on tape. Later on, all opinions are analysed and the most promising culled out. The rationale of brain-storming may be defined as follows: the greater the number of ideas advanced, the greater the proportion of those having merit.

The *cross-examination* technique consists in that a group of experts cross-examines the members one by one in turn. The questions are formulated in advance to a predetermined algorithm so as to cover a wide range of problems. The technique utilizes man's ability to establish relationships between things analytically and, often, intuitively, this ability being stimulated by cross-examination.

The *court-of-justice technique* is based on the fact that the court procedure perfected in the course of centuries is extremely efficient in bringing to light all factors, however insignificant they might seem, which may be conducive to truth. The technique works as follows. One of the group members is nominated the advocate, another the attorney, and both select witnesses from among appropriate experts. The remaining members act as judges, and the problem at hand as the accused. Then everything goes on as in a real lawsuit.

Questionnaires are advantageous in that they are amenable to mathematical analysis. Most often, use is made of the Delphi technique first applied by the Rand Corporation of the United States to a wide range of problems: armament build-up, the choice of weapon systems, power-system expansion, etc. The technique reduces to the following. Appropriately formulated questionnaires are circulated among, say, 100 to 200 experts who are invited to give their views and estimates. Processed statistically, these estimates suggest the courses of action to be followed and the objectives to be met. At the conclusion, a final decision is made on how to tackle the problem.

Some examples of forecasting deserve a closer examination. Suppose we are to assign measures to n items on the basis of N criteria. Each of these criteria may lead to any one of the following measures:

(a) a binary measure of the YES – NO type, stating whether or not a given item has a certain quality;

(b) a discrete measure out of a specified finite set of measures;

(c) a numeric measure which takes any value within a specified interval.

Measures (a) and (b) may be qualitative (good — bad; excellent, good, satisfactory, etc.).

Suppose further that each item examined can be assigned measures on the basis of all the criteria. Then the item can be visualized as a point in an N-ary space of criteria.

The next step is to identify which of the items is the best, second-best, and so on, that is, to order the items. In many cases, this step may be preceded by what may be called an aggregation of measures, when the various measures based on the particular criteria are combined into an overall measure.

Utility theory. The problem taken up in the previous example may sometimes be tackled by the theory of utility as part of the expert-opinion method. Although this approach has long been used and many of its points are widely accepted, its practical application is rather limited and the solutions it yields are not always the best.

By this method, one seeks to prove the existence of a single-valued scalar function $U(x)$ of a vector x, and measures against criteria which enable the items under consideration to be well ordered. The function $U(x)$ assigns a scalar measure to each item, after which the ordering problem is solved in an obvious way.

There is a prevailing body of opinion that the function $U(x)$ exists irrespective of the way in which a decision may be made. In contrast, opinions and theories differ on the form of this function, that is, on the manner in which the *measures by different criteria can be combined into an aggregate measure.*

Recently, a theory of additive utility has gained some currency. It reduces to the following. A scale of measures is associated with each criterion, and the items are assigned appropriate measures from these scales. The various scales are related by factors (usually called criterion weights). Then the function $U(x)$ is found as the sum of the measures multiplied by the respective weights (the method of weighted sums).

The theory has met with justified criticism. The most controversial point is that the deficiency in the quality associated with one criterion may be compensated for without bound by an improvement in the quality associated with some other criterion. This may happen because in an aggregated measure the individual criteria are indistinguishable, and the same value of $U(x)$ may well apply to a variety of combinations of the individual criteria. Indeed, some of these combinations may be absurd from the view-point of common sense.

In an attempt to circumvent this limitation, quite a number of other methods for determining the function $U(x)$ have been proposed. Mention may be made of the following:

(1) the use of nonlinear coefficients relating the criterion scales and based on the dependence of these coefficients on the measures themselves;

(2) the use of the function $U(x)$ strongly varying with measures by individual criteria;

(3) the use of only 'bad' measures against individual criteria ('good' measures are ignored).

Each of these techniques has merits and demerits of its own, and each should preferably be used in a particular area, which fact is examined at length by various authors. It is only to be noted here that the first method involves a considerable amount of computational work, if criteria are many. The second method does not avoid the limitation of the weighted sum method (that is, deficiency compensation). The third method fails to take account of all information available.

For all their merits and demerits, the various methods in use today are only means and tools for deriving additional information and cannot affect decision-making as such. This fact stems directly from the nature of the matter.

All methods based on aggregation of measures seek to give a rational solution in an area where formalization is difficult, if at all possible. The point is that many criteria are rather subjective; also the relationship between the best and worst may be different (linear or nonlinear); whether or not a given item having a given combination of measures is superior to that having any other combination may depend not so much on the measures as on the population of items considered. Moreover, the superiority based on the relations (weights) selected subjectively may sometimes run counter to the intuition of the person who has chosen these relations (weights).

Thus, these matters are not purely mathematical or economic; they have also a psychological facet which must be taken in account in the search for new methods. The understanding of these limitations has already led to a number of modified methods for aggregation of measures. A distinction of these methods is that the measures are aggregated with the participation of the person who has stated the problem and will use the solution (that is, the actual decision-maker).

The next group of methods we are going to discuss applies mathematical programming to systems involving a multitude of criteria. This group includes the Electra method which brings out unambigous relationships between the items or events considered by an analytical procedure. The events whose comparison entails considerable difficulty are declared uncomparable at a given stage. If their comparison is essential for decision-making, the following two alternatives are possible:

(1) the problem is passed on to the decision-maker; or

(2) the comparison is carried out analytically and suffers from drawbacks similar to those in the method of weighted sums. The difference is that the items are now compared sequentially, with the most unambigous relationships being established first and those involving a degree of arbitrariness

being deferred until the last stage. From what has been found in the various stages, the decision-maker is in a position to choose the most unbiased policy.

Before we take up the above methods in greater detail, it is worth while dwelling on the ways and means of securing appropriate information.

Acquisition of expert information prior to and at the time of decision-making. The acquisition of objective information is the most essential requirement for any control system to make sound decisions. Since in organizational structures information mainly comes from or through people, any management is faced with a very important task in its own right — that of extracting objective information from experts' subjective opinions. A reliable expert opinion on any matter cannot be obtained without assistance from people who know scientifically how to set up expert groups, guide their work, and cull out a final decision from their judgements.

The expert opinion procedure is an extremely important and sophisticated technique which can be successful only if the experts are fully aware of their responsibilities and apply all of their creative thinking. The crucial aspect of the technique is that judgement has to be passed on a fairly broad problem on the basis of what the experts have spoken on its particular aspects.

The rationale of deriving valid information from experts' opinions in organizational management systems is as follows. Each expert is presented a hypothesis, H, and he participates in the selection of indicators, E, which are logically related to it. On the basis of these indicators, the group arrives at an a posteriori hypothesis probability, $P(H/E)$, that is, at an estimate of the probability that the hypothesis will materialize, subject to the occurrence of the indicators E. Thus, $P(H/E)$ is a measure of the likelihood of H on the basis of the assumed truth of E.

A very simple case may serve as an example. Let E have the form of a statistical record which states that m out of n items examined have a property K, and the hypothesis H assigns this property to an object not examined yet. In the circumstances, the a posteriori probability of the hypothesis H on the basis of E is defined as the frequency m/n.

Where E cannot be presented in a simple statistical formulation and H is not a statement like the one described above, the hypothesis $P(H/E)$ cannot be defined uniquely, and it is only possible to define an interval of values within which $P(H/E)$ may be found. Now the a posteriori probability of the hypothesis is the probability estimate ventured by each expert.

In this case, the procedure is as follows. An expert is presented a hypothesis the truth of which is yet to be established, and information which together with the expert's presumed knowledge forms the indicators E characterizing the hypothesis. The expert states his personal estimate in terms of probabilities. Hence, an expert is expected to comply with the following requirements:

personal evaluations must be relatively invariant with time with E held unchanged;

new indicators should affect personal evaluations in the correct direction.

In simple cases where $P(H/E)$ can be computed on the basis of the known E, personal evaluation should agree with $P(H/E)$.

The expert taking part in the procedure must be well versed in his particular field, and his forecasts must therefore be successful for a long time period.

Where several experts (a panel) are working together, they return a series of personal evaluations, and the task of combining them arises. This can be done in any one of several ways:

a favourite expert is chosen and only his opinion is taken into consideration;

the probability estimates of the individual experts are averaged by determining their median or mean weighted value (in which case different weights may be assigned to the experts);

the experts arrive at a consensus through a discussion (the limitations of this approach have already been stated).

It is important, however, to remember that experts give, strictly speaking, likelihoods and not probabilities. The calculation of likelihood has been widely described in the literature. It is stressed that although some of the rules borrowed from the theory of probability do apply to likelihood estimation, likelihoods cannot be viewed as being numbers, because they only form a partially ordered set.

V. M. Glushkov's method. In brief, this method is as follows. Prior to the actual evaluation, a panel of experts:

formulates a set of so-called closing events, that is the events for which the panel seeks to determine the probability of occurrence and realization;

an ordered listing is drawn up of the events that (as the panel assumes a priori) must or may turn up in a multitude of situations which will enable (in accordance with the starting assumption) a judgement to be passed on whether the sought objective is realizable (whether or not there is a solution to the problem);

on the basis of the events included in the listing, a tentative list of experts is drawn up and each expert is assigned an a priori weight for each type of events included in the listing.

The first step in the actual procedure is to make known to each expert the events for which he is to predict the time and probability of occurrence. The experts qualifying for this job are those whose weights for the events involved exceed a predetermined level.

Let s be the event, the probability and time of occurrence of which are to be estimated. In general form, the condition P for the event s to occur may be expressed as an arbitrary logical function $f(s_1, s_2, \ldots, s_k)$ of some independent (in the expert's opinion) events $s_1, s_2, \ldots, s_k$.

Then each expert estimates the conditional probability $P_p(s)$, that is, the probability of an event s occurring subject to the condition P, and the most probable value of time, $T_p(s)$, that may elapse between the instant

when the condition P is satisfied and the instant when the event s occurs (if it does so at all). In doing so, the expert may (and, indeed, he is encouraged to) give an estimate of the unconditional probability of the event s occurring, and of the total time period from the present time to the instant of occurrence. In this case, the condition P reduces to an identically true event (that is, the events $s_1, s_2, \ldots, s_k$ form an empty set).

The questionnaires returned by the experts form a basis for a network of events, with each expert's estimates of $P_p(s)$ and $T_p(s)$ entered in a suitable network model. To simplify the processing of questionnaires, appropriate modifications are made in the network. Since the experts inevitably introduce new events, these are sent to other experts for evaluation, and this procedure is repeated until the network is stabilized. The experts taking part in the previous round also participate in the next. In this way, the list of events, now extending from s_1 through s_n, is finally stabilized, as is the list of experts, and the network assumes its final form suitable for further mathematical processing.

The distribution of the absolute time of occurrence of any event s within the network under consideration is characterized by the vector of probabilities $[P_1(s), P_2(s), \ldots, P_k(s), P_\infty(s)]$, such that $P_i(s)$ is an estimate of the probability that an event s will occur before the instant of time t_i. Notably, $P_\infty = P$ is an estimate of the unconditional probability that such an event will occur within an infinite time period.

An overall estimate of P_i is derived by averaging the estimates returned by the individual experts. The averaging (or smoothing) algorithm is based on the weights assigned to the experts. Averaged estimates are obtained sequentially, in moving from a previous event to the next.

Rank correlation. Ranking. A population of items may be arranged in descending or ascending order of some non-quantifiable property. The ordering thus produced is called a ranking; from it, we can determine the rank, x_i ($i=1, 2, \ldots, n$) of any item within the ordered arrangement.

Concordance between the opinions of two experts who have ranked n items (or n main directions of a development) can be evaluated in terms of coefficients of rank correlation.

Assume that two experts have ranked n entities in accordance with their personal preferences as follows: the first expert, $x_1, x_2, \ldots, x_n$; the second expert, $y_1, y_2, \ldots, y_n$, where x_1 through x_n and y_1 through y_n are natural numbers from 1 to n arranged arbitrarily. The ranks may be regarded as numbers, and ranking as the assignments of numbers to the events ranked.

Consider the entities numbered i and j, where $j=1, 2, \ldots, n$. The first expert has given them the ranks x_i and x_j, and the second, the ranks y_i and y_j, respectively. Let the relation between x_i and x_j be a_{ij}, and that between y_i and y_j be b_{ij}. We also let

$$a_{ij} = -a_{ji}, \qquad a_{ij} = 0 \quad \text{for} \quad i=j$$
$$b_{ij} = -b_{ji}, \qquad b_{ij} = 0 \quad \text{for} \quad i=j$$

Then the coefficient of rank correlation will be

$$\Gamma=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}b_{ij}}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2\sum_{i=1}^{n}\sum_{j=1}^{n}b_{ij}^2}}$$

As is seen, it may fall between $+1$ and -1. When $\Gamma=1$, this is an indication that the ranking proposed by the first expert is fully in agreement with that proposed by the second. When $\Gamma=0$, no concordance exists between the two rankings. When $\Gamma=-1$, the first expert's ranking is exactly opposite to that given by the second.

Different expressions can be derived for the coefficient of rank correlation, depending on the manner in which a_{ij} and b_{ij} are defined.

In order to test the significance of the coefficient of rank correlation, that is, the degree of concordance between two experts who have ranked n events, it is necessary to assess the significance of the result.

Let it be known that there is no concordance between the rankings offered by two experts and also that the first expert's ranking is a sequence of natural numbers. Then the ranking proposed by the second expert can be examined by correlating it with the natural series. Since no concordance has been assumed to exist between the two rankings, all rankings that the second expert could propose are equi-probable.

For each ranking there is a specific value of the coefficient of rank correlation. Let f designate the frequency at which a particular value of the coefficient can occur. In the limit of high values of n, the frequency distribution is symmetrical and tends to the normal distribution

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-x^2/2\sigma^2\right)$$

with a standard deviation

$$\sigma=\sqrt{1/(n-1)}$$

At $n>10$, the frequency distribution is approximated by the normal distribution curve with sufficient accuracy.

Where more than two experts participate in the procedure, the so-called *coefficient of concordance*, W, is introduced. As its name implies, the coefficient gives a measure of concordance between the experts on the ranking of entities by a chosen quality. The coefficient of concordance falls between 0 and 1. When $W=0$, there is no concordance between the rankings, that is, no consensus among the experts. When $W=1$, this is an indication that all experts have ranked the entities under considerations in precisely the same manner, that is, all experts share the same view on the ranking of the entities.

Mention should also be made of the *preferential ordering method.* In contrast to rank correlation, this method seeks to find a ranking which would

be closest to that of each expert, rather than to determine the degree of concordance between them. The concordance between two rankings A and B is determined by the separation between them

$$d(A, B)=\frac{1}{2}\sum_{ij}|a_{ij}-b_{ij}|$$

where a_{ij} and b_{ij} are found for the first and second rankings in the same manner as in the rank correlation method.

The average ranking may be found in any one of several ways. If there are several ordered arrangements A_1, A_2, ..., A_m (not necessarily all of them different), the ranking B, for which

$$M=\sum_{i=1}^{m} d(A_i, B)$$

is a minimum, is called the *median*, and that, for which

$$S=\sum_{i=1}^{m} d(A_i, B)^2$$

is a minimum, is called the *average*. There is a certain difference between the median and the average. The median reflects the opinion of the majority; where no majority exists it is not unique and may apply to several ranking separations, all of them giving the same minimum. With averages, they are all taken to be of the same significance, if the experts have given different ranks to different entities.

1.3. Modelling

Models intended for use in forecasting markedly differ from those employed in system expansion planning. Above all, the former are less detailed. Yet, there may be quite a number of features common to models of both types.

Modelling is a method of study in which the process or phenomenon of interest is replaced by another process or phenomenon not necessarily analogous physically, but more convenient for analysis. In mathematical modelling, a particular process is represented, or rather abstracted, by a system of mathematical equations which describe logical relationships that can help in the investigation of the process.

As is seen, modelling is a very broad concept. It may well apply to a wide range of problems ranging from the inception of an idea (Fig. 1.10*a*), through feasibility studies, realization of a project, development, engineering to production (Fig. 1.10*b*).

With regard to forecasting, several aspects must be taken in account in modelling. Here, emphasis is placed on mental experimentation based on the internal logic of developments, and conservative use of historical analogies.

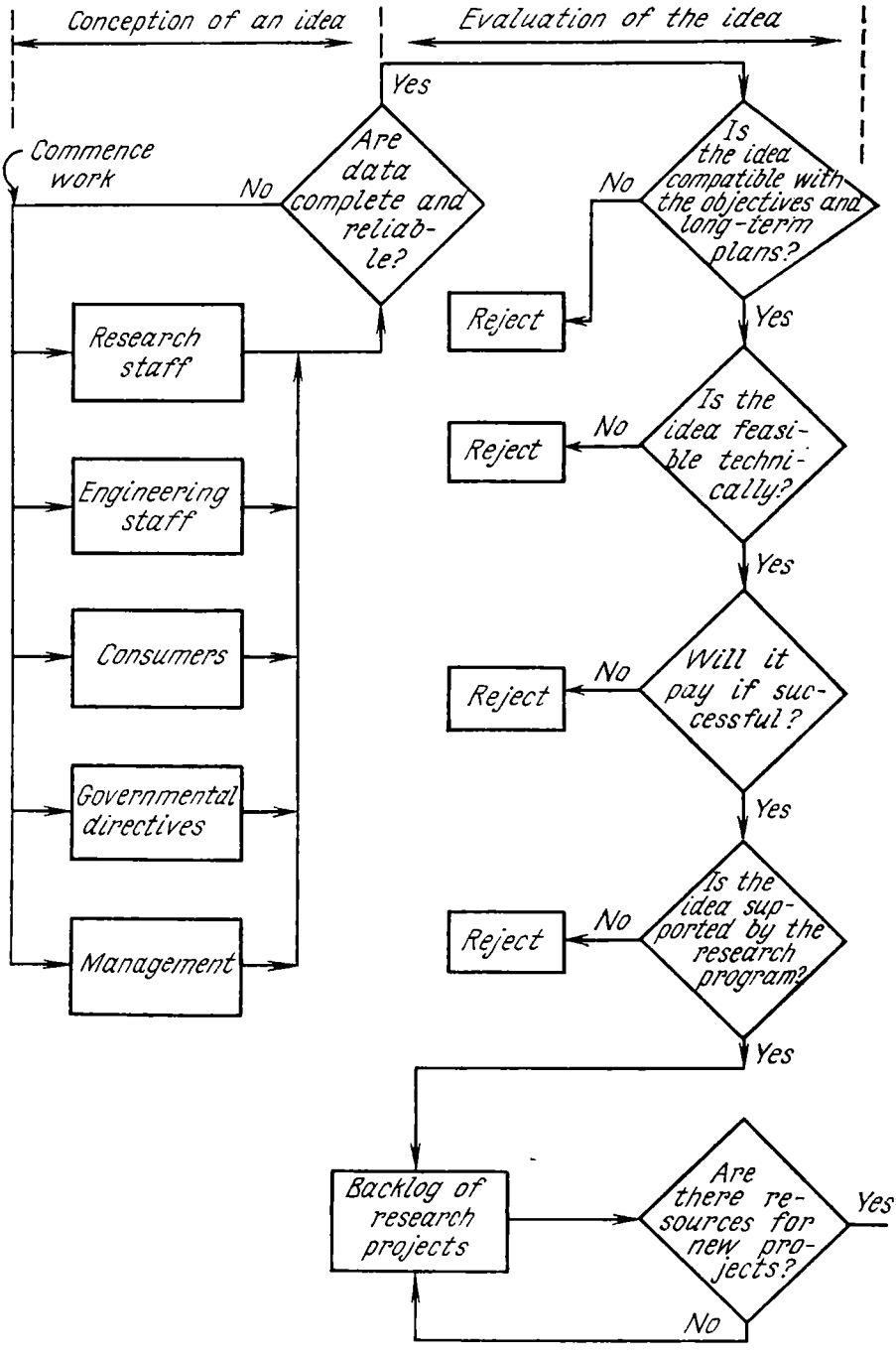

Conception of an idea
Evaluation of the idea
Commence work
Are data complete and reliable?
No
Yes
Is the idea compatible with the objectives and long-term plans?
No
Yes
Reject
Research staff
Engineering staff
Consumers
Governmental directives
Management
Is the idea feasible technically?
No
Yes
Reject
Will it pay if successful?
No
Yes
Reject
Is the idea supported by the research program?
No
Yes
Reject
Backlog of research projects
Are there resources for new projects?
Yes
No

Fig. 1.10*a*

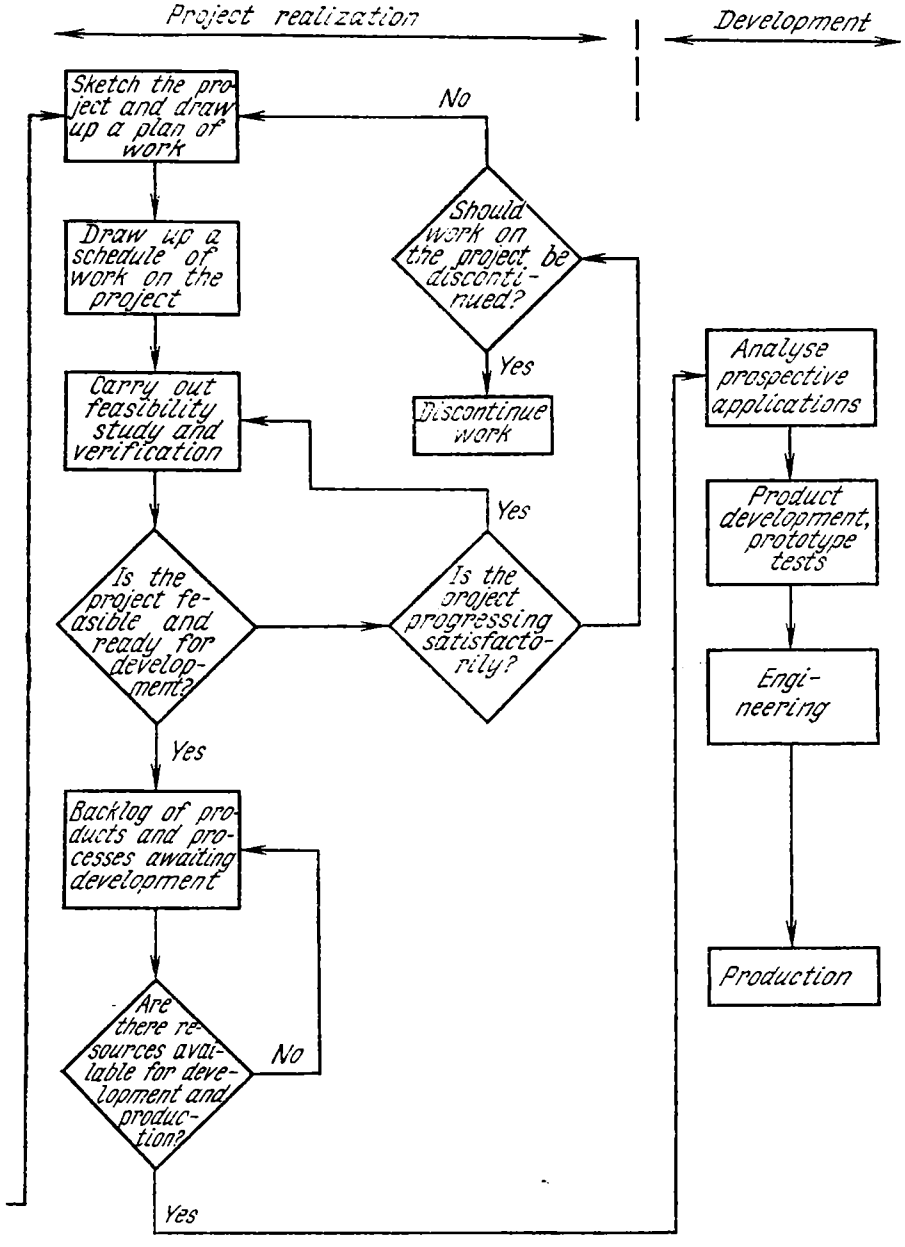

Project realization
Development
Sketch the project and draw up a plan of work
No
Should work on the project be discontinued?
Draw up a schedule of work on the project
Yes
Discontinue work
Carry out feasibility study and verification
Yes
Is the project feasible and ready for development?
Is the project progressing satisfactorily?
Yes
Backlog of products and processes awaiting development
Are there resources available for development and production?
No
Yes
Analyse prospective applications
Product development, prototype tests
Engineering
Production

Fig. 1.10*b*

As a rule, logical models are used to represent a comprehensive picture of progress in science or technology, and include the formulated economic, political and other objectives of a nation, a description of relevant scientific and technological achievements, and a characterization of resources and demands. This descriptive document may be called a *forecast scenario*, and the forecasting method based on it, the *scenario method.*

In forecasting, significant importance is attached to mathematical models, especially to those which approximate the future of a process by a known statistical distribution. For example, the life cycle of mass-scale technical means can be approximated with sufficient accuracy by the normal (Gaussian) distribution.

Much can be done in forecasting through the use of trend curves. In an elementary case, this is an exponential curve defined as

$$S_t = S_0 \exp(kt)$$

In cases where an approximation is sought for a quantity the growth of which is bounded by fundamental causes, such as the number of scientists, the number of projects realized, etc., resort may be made to an expression of the form

$$S_t = b/[1 + S_0 \exp(-kbt)]$$

where b is the maximum likely value.

A generalized model for a group of related trend curves may be built as follows. The effectiveness with which the means M can be utilized is written as a function, $f(M)$. The payoff accruing from improvements in a given type of technical means is expressed in terms of the parameter P defined as

$$P = \varphi[f(M)]$$

for which the limit, P_{lim}, is set such that

$$P < P_{lim}$$

As P approaches P_{lim}, the increment in the payoff parameter, dP, decreases

$$dP = (P_{lim} - P)\, dM$$

Also, dP depends on the previous effort

$$dP = (P_{lim} - P) f(M)(dM/dt)\, dt = (P_{lim} - P) g(t)\, dt$$

The solution of the above differential equation may be written as

$$P = P_{lim} - (P_{lim} - P_0) \exp\left[-\int_0^t g(t)\, dt\right]$$

For different design or input data, a set or family of trend curves is usually obtained.

Mathematical models also include those in the form of functional hierarchical networks. These models provide an insight into the complexity of

the problems awaiting solution, the risks associated with forecasts, and the probability of occurrence of the events forecasted. A specific form is a *task tree* which enables a task to be decomposed into subtasks readily analysable.

A specific role in forecasting is assigned to information *modelling*. This approach is used in the multi-dimensional analysis of patent applications with a view to determining what new they offer in the field of technology. Techniques are being developed to pool information on the world's technical know-how from the patents already granted. One of such methods breaks down the work into three stages as follows:

(1) each of a collection of patents is assessed for completeness (engineering significance and economic payoff);

(2) the aggregate significance indicators of competing inventions are compared to determine which of them is more promising;

(3) a measure is put on the overall level of patents in a given section, and the trend of development is stated in verbal form.

Measures are stated in points. The completeness factor, T, of an individual invention is stated as the ratio of the sum of measures, q, actually given, to the maximum likely one, Q

$$T=q/Q$$

The sum of completeness factors T over a group of n patents is called the weighted patent number M

$$M=\sum_{1}^{n} T$$

The ratio of the weighted patent number $M(\tau)$ to the nominal one, $N(\tau)$, over the same number of years gives a generalized completeness factor

$$T_{gen}=\int_0^{\tau} M(\tau)\,d\tau / \int_0^{\tau} N(\tau)\,d\tau$$

$$0.2 \leqslant T_{gen} \leqslant 1.0$$

A given trend is promising if $T_{gen}=0.8$ to 1.0. If $T=0.2$ to 0.4, there is no sense in continuing the development.

Recently, means have been developed which automatically establish indexes of *relevance* between various publications. The relevance index method has proved to be a very powerful tool in analysing and forecasting development trends in science. Basically, the method produces a relevance network which models the historical and logical relationships underlying a given development.

An example of modelling is the PATTERN technique (which is the acronym of Planning Assistance Through Technical Evaluation of Relevance Number) which has originated in the United States. It involves a number of separate operations, each dealing with several alternatives. The scheme is based on expert opinions and a hierarchical decomposition of

tasks: the overall task at hand is decomposed into subtasks, each of which is further divided into sub-subtasks, etc. At each level, experts rank the subtasks and assign them numerical weights to establish priorities. On the basis of the priorities thus found, money is allocated to each subtask. The decomposition is continued until any one is unable to bring out a specific task. Various factors of which there may be 10 to 20 are examined in each task. An important advantage of the scheme is that it can forecast world developments (economic, political, etc.) by 'acting out' a scenario. For this purpose, a scenario of, say, political developments is written, in which consideration is given to the international situation, a balance of political forces, the state of world economy, the quantity of food needed, the likely theatre of war operations, strategic intelligence, etc., that is, to every thing that might have any bearing on the likely outcomes. From reference to the relevance tree, the experts then decide which action should be planned for which relevance, and assign a relevance number to each action. In planning for 1980, for example, consideration has been given to several thousand technical tasks important to decision-making at the next higher level. The following relevance numbers have been assigned: the country's survival, 0.6; demonstration of good will, 0.3; public opinion, 0.1.

Thus, an analysis of the 'state of nature' leads to an evaluation of the problem to be solved and the probability of occurrence of particular events. In other words, the scheme seeks to fit a playing strategy to a situation, which is a particular use of the theory of games, an integral part of operations research.

Operations research has emerged as a result of concentration on a particular class of management or executive problems. Originally those were military problems of strategy and tactics for the most effective utilization of very limited resources. As already noted, it also includes the theory of games specifically developed to deal with conflicting situations and to provide solutions on a mathematical basis. Games are classed into *open*, where man is participating and the human factor has to be evaluated; and *closed* where the problem of interest is formalized, presented as a mathematical program and solved on a digital computer.

Sometimes, games are also classed into *war* games where problems of strategy and weapons are dealt with, and *business* games dealing with competitive problems in the conditions of capitalist society.

All of these schemes follow basically the same pattern: a relevance tree is drawn up, relevance numbers are assigned, running estimates are derived, mutual utility and prospects of development are assessed, and a scenario is worked out.

There exist widely differing opinions on how to assess alternative forecasts. For example, Academician V. A. Trapeznikov proposes that the effectiveness of forecasts be evaluated in terms of the coefficient of payoff on capital investments. He defines his payoff coefficient as the ratio of ΔP to ΔP_m, where ΔP is the growth due to technical progress and ΔP_m is the increase in expenses caused by the expansion of technical means used.

Professor B. G. Kuznetsov proposes that the effectiveness of alternative forecasts be assessed in terms of derivatives. For a forecast to be effective, he argues, the first derivative dP/dt should be greater than zero, where P is the productivity associated with a given alternative. Obviously, the second derivative, d^2P/dt^2, must likewise be greater than zero; it describes the effectiveness of technical progress and applied sciences. The third derivative, d^3P/dt^3, describing the effectiveness of fundamental research, should be greater than zero, too. In Kuznetsov's opinion, for a structure to be optimal it is not enough for productivity to be a maximum; what actually matters is to maximize the product

$$P(dP/dt)(d^2P/dt^2)(d^3P/dt^3)=\max$$

Progress in science and technology and the contributing factors may be assessed in terms of the basic parameters listed in Table 1.1.

It is to be noted that the methods just mentioned are not generally accepted, and they have been cited only as examples of likely criteria of effectiveness.

With respect to power-industry problems, our concern will be with three gradations of tasks: long-term forecasts for 15 to 30 years in advance, short-term forecasts for 5 to 15 years ahead, and current plans.

For the purpose of planning, two basic types of models are used, optimization and estimation models. An *optimization model* yields a mathematical solution in the form of a minimum or a maximum. An *estimation model*

Table 1.1

Parameters of Scientific and Technological Potential

Contributing factor	Parameters of scientific and technological potential			
Supply of trained personnel	Total number	Qualification mix	Age mix	Resources and reserves
Supply of scientific information	Backlog of own ideas and theories	Knowledge of world developments	—	—
Material basis	Budget allocations	Supply of appropriate instrumentation and physical facilities	State of art and potentialities of industry	—
Organization of science	Strategy in development of science (criteria in choice of trends)	Level of optimality in scientific potential	—	—

produces an estimate of each alternative submitted by the designer. Each of these models uses different programming techniques. In most cases, resort is made to *linear* and *nonlinear* programming.

As a further step, models may be classed into global and unitized. A global model leads directly to an assessment of an optimum alternative as a whole; a unitized model does so in parts only, and an overall estimate is obtained by combining the partial estimates by iteration. Here, too, programming may be linear or nonlinear, continuous or discrete (integer). It may be noted that linear programming includes a wide range of specific methods and techniques, such as the simplex algorithm, or simplex method, Gomory's method, transportation problems, etc. Nonlinear programming includes the gradient method, dynamic programming, the bound and branch method (see Chap. 2).

The tasks of forecasting (stage I) and of planning (stage II) are listed in the chart of Fig. 1.11.

Economic mathematical models for electric-power systems. Model-building for the purpose of technical and economic analysis of an electric-power system always involves three basic steps:

(1) Establishing criteria of optimality.

(2) Constructing a mathematical model of the system.

(3) The choice of a solution method*.

Criteria of optimality. The overall criterion of optimality for complex technical and economic systems, either being built or already operating in socialist society, can best be stated as "the most effective and rational utilization of material wealth and manpower to satisfy the growing demands of society members".** Formalization of this criteria in optimization problems leads to specific criteria or an equivalent overall criterion. Among the former are well-known economic criteria, those associated with social and demographic factors, the effects of engineering systems on the biosphere, criteria reflecting the progress of technology, rates of growth of labour productivity, etc.

Of course, if in a complex cybernetic system developing in a purposeful manner and governed by deep internal, often stochastic relationships, we could have complete information about all of these relationships and assess all controlling factors in terms of money, it would be possible to reduce everything to a single criterion — the minimum discounted (weighted) cost. Unfortunately, however attractive it might seem, we are not yet in a position to combine in a single criterion, such as the minimum discounted cost, many of the factors, for example, demographic or impact on the biosphere, because of lack of information at present.

Even if we selected the minimum discounted cost as the only indicator, an optimum solution could not be visualized as a single point in a multi-

* See Venikov V. A., Arzamastsev D. A. *On Model Building for Electric Power Systems.* USSR Academy of Sciences. "Energiya i transport", 1970, No. 2.

** Program of the Communist Party of the Soviet Union, "Gospolitizdat", 1962.

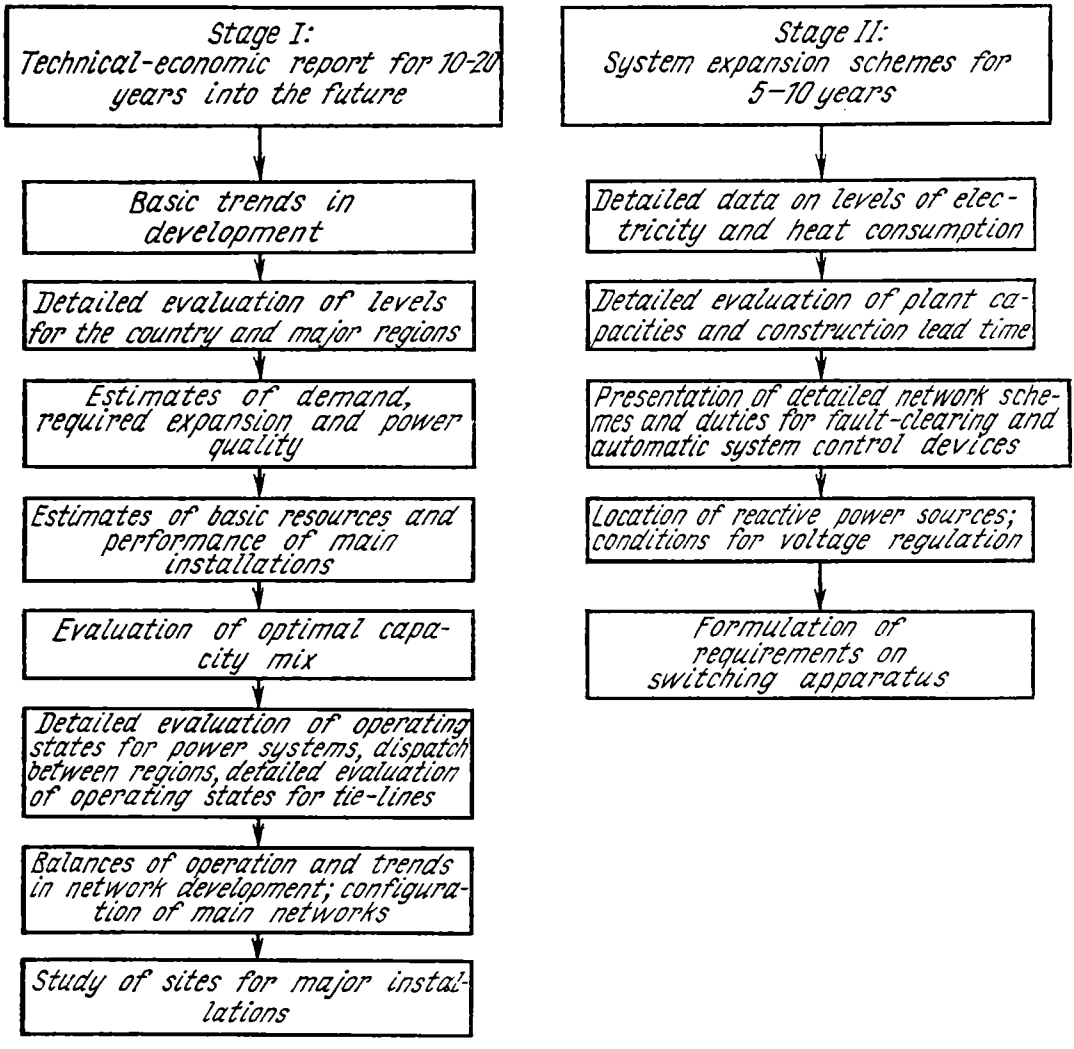

Stage I:
Technical-economic report for 10-20 years into the future
Basic trends in development
Detailed evaluation of levels for the country and major regions
Estimates of demand, required expansion and power quality
Estimates of basic resources and performance of main installations
Evaluation of optimal capacity mix
Detailed evaluation of operating states for power systems, dispatch between regions, detailed evaluation of operating states for tie-lines
Balances of operation and trends in network development; configuration of main networks
Study of sites for major installations
Stage II:
System expansion schemes for 5–10 years
Detailed data on levels of electricity and heat consumption
Detailed evaluation of plant capacities and construction lead time
Presentation of detailed network schemes and duties for fault-clearing and automatic system control devices
Location of reactive power sources; conditions for voltage regulation
Formulation of requirements on switching apparatus

Fig. 1.11

dimensional field reflecting relevant factors all the same, because input data would be incomplete and uncertain.

In other words, formalized optimization problems usually treated as having a single criterion necessitate consideration of several criteria. Obviously, this also applies to electric-power systems where, apart from the minimum discounted cost, other criteria apply, selected according to the properties and application of a particular power system interoperating with other systems essential to the existence and development of socialist society as a whole.

Naturally, a practical optimization of system expansion should seek to satisfy both the main criterion and the necessary additional, or partial, criteria.

From the foregoing it follows that in the optimization of system expansion we may, and indeed must, speak of an *expansion strategy* associated

with a multiplicity of equi-optimal courses of action. In the general case, it would be wrong to expect that a single optimum plan exists as some singular point in the space of strategies.

It is vital to narrow down this multiplicity of equi-optimal courses of action. However, whether this narrowing-down is possible is markedly dependent on several circumstances the most important of which are the accuracy of optimality criteria and the validity of assumptions usually made in stating objective functions. Study in this field is far from completion, which is borne out by many examples, such as the economy criterion, one of the most commonly used criteria.

From the view-point of economics, the optimization of production expansion reduces to improving the relative effectiveness of labour. Here, one of two economy criteria may be used:

(1) to maximize the economic payoff from the aggregate expenditure of material and labour resources allocated at the beginning of the time period under consideration;

(2) to minimize the labour expenditure needed to satisfy the specified 'identity of effect' condition, subject to the constraints imposed on the multiplicity of valid courses of action. In this context, 'identity of efffect' means that the output of products must meet appropriately specified requirements for the quantity, quality, time and place of production and consumption.

Each of the two criteria is a partial one as it is specific to only one aspect of the overall optimality criterion. In a sense, they are equivalent, for where the second applies to local constituent systems the first applies to the national economy as a whole.

In this book, the first criterion is not examined as it lies outside the scope of our discussion. Generally, it involves the consideration of the present and future for generations to come, and also some other problems, either debatable or defying formalization.

The second criterion is predominantly used in specific problems such as the optimization of systems local with regard to the national economy, although some of them are complex systems in the cybernetic sense of the word. Mathematically, this criterion requires to minimize an economic functional we have called the discounted or weighted cost. As has been noted more than once, this is the principal criterion in practical planning. At present, there is no generally accepted opinion on many aspects involved in the derivation of the discounted cost. Among other things, it is still unclear whether a single or several standard coefficients of effectiveness should be used, how they should be determined, how account must be taken of the time factor, what time period should be adopted as a reference in optimization, how the balance cost of installations should be found and the dynamic behaviour of a system should be accounted for.

Constructing a mathematical model of a system. In constructing a mathematical model, the following steps are involved: (a) select a class of model;

(b) take account of the probabilistic character of available information; and (c) formulate requirements for the precision of the model.

As already noted, estimation and optimization models may be used in planning for the optimal expansion of power systems or of their functionally or geographically separated elements.

Estimation models are realized by running a predetermined set of alternatives on digital computers, using appropriate algorithms and computational programs. Among the advantages of these models are: (1) the specific features of each projected power system can be investigated in detail in each alternative design proposed; (2) they are insensitive to changes in system variables and the functional relations between them; (3) each alternative design can conveniently be tested for compliance with reliability requirements and refined to incorporate the necessary range of reliability characteristics.

A major disadvantage of estimation models is that only a limited number of alternative system designs can be compared. This is fraught with unwarrantable reliance on subjective judgement and, as a consequence, with the risk of leaving the optimization domain and giving preference to a non-optimal strategy. This is why estimation models necessitate a rigorous motivation for the choice of a particular alternative.

This limitation has mainly been responsible for resort to optimization models. As their name implies, they seek to choose an optimum plan from among a multitude of feasible or viable plans, which is a major advantage of this type of models.

Although, theoretically, optimization models are applicable to a very wide range of problems, there are appreciable difficulties in their realization. For one thing, in many cases optimization models have to be simplified to the extreme, and they cannot therefore offer the advantages of estimation models. For another, power-system problems, when stated in rigorous form, are often nonlinear, may have several extrema, and involve multiple variables in integer (discrete) form. This adds to the complexity of building optimization models and, especially, of the computational techniques involved.

For some time past, there has been a tendency to combine both types by incorporating some optimization blocks within an optimization model of system expansion. With such an arrangement, the functional defined on the complete set of feasible alternative plans can be individually minimized for each partial problem. This approach has definite advantages although it cannot resolve the principal limitation of estimation models.

There is another alternative which has so far been given scanty theoretical treatment — the two types of models can be used sequentially. An optimization model could be used for a preliminary comparison of the alternatives to find those approaching an optimum solution (in terms of the objective function, because they can markedly differ in the variables involved). Then an estimation model could be used for a comparison of the remaining alternatives. This approach could utilize the advantages of

each type and minimize the limitations of both. This could be especially advantageous where problems may have several minima or maxima or where it is essential to take into account the likely variations in the 'identity of effect'. A general classification of mathematical models appears in Fig. 1.12.

Whatever method is chosen for problem solving, it is important that it should be able to handle information which is probabilistic in character. If we trace the approaches that may be taken in solving any general optimization problem, we shall note that the choice is usually between a deterministic and a stochastic approach.

A *deterministic approach* assumes that the input (design) data uniquely describe external factors and the state of the system and that a single optimum solution can therefore be found. Deterministic models are relatively simple and lend themselves to solving readily.

Actually, however, not all practical time-varying systems and external factors can be uniquely defined in advance. This is where a *stochastic approach* has to be used, which assumes that only some part of input data may be deterministic, while the remainder demonstrates stochastic characteristics because it includes random variables and functions thereof. In such a case, too, the problem is solvable, provided we know the relevant probabilities. It is to be noted that the longer the time period over which a system is studied, the less we can know about external influences and about the likely qualitative changes that may occur in the system in the future. In other words, the information we get becomes increasingly more uncertain.

In the circumstances, it is important to take account of the decrease in the certainty of information relating to the remote future as we extend the time period of optimization. This is true of not only probabilistic information and probabilistic models but also of cases where deterministic models utilize data presented as the expectations of random quantities or functions.

According as we choose a probabilistic or a deterministic approach, we can single out two groups of problems. One will encompass *long-term forecasts* intended to outline general trends of system development. Installation parameters and start-up dates are not the main concern in such cases, although these points may be of interest, too, especially with regard to critical units defining the basic system structure. Because of the long optimization period involved, relevant data are mostly probabilistic and partly with a good measure of uncertainty (whatever deterministic information may be present is of minor consequence). Obviously, the only choice is to use some kind of a probabilistic model. However, this approach as such does not do away with the uncertainty of input data; this uncertainty is in fact carried over into the solution and would increase as the optimization period is extended farther into the future.

The other group of problems are those concerned with *planning* and intended to fix specific system variables and startup dates.

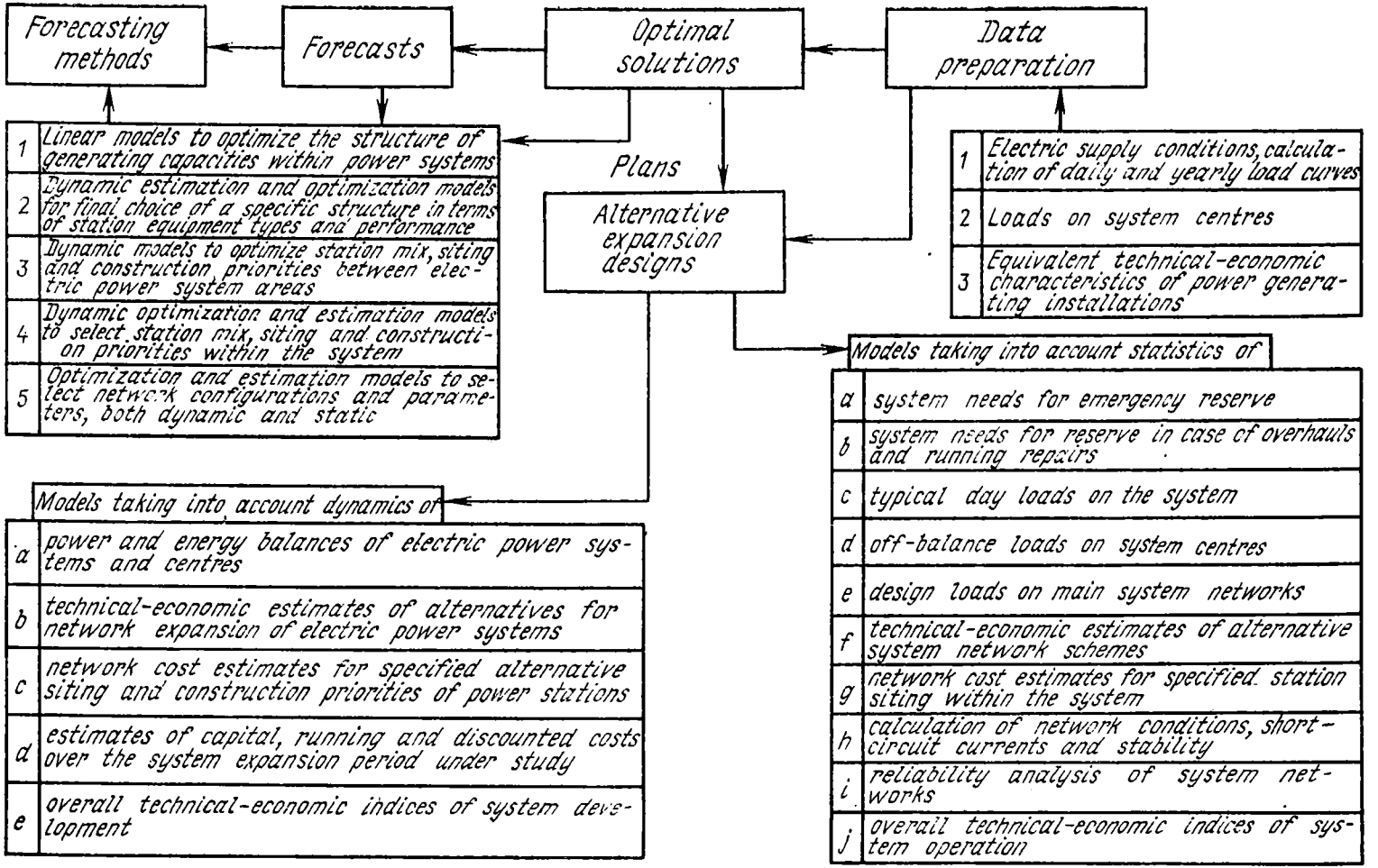

Forecasting methods
Forecasts
Optimal solutions
Data preparation
1 Linear models to optimize the structure of generating capacities within power systems
2 Dynamic estimation and optimization models for final choice of a specific structure in terms of station equipment types and performance
3 Dynamic models to optimize station mix, siting and construction priorities between electric power system areas
4 Dynamic optimization and estimation models to select station mix, siting and construction priorities within the system
5 Optimization and estimation models to select network configurations and parameters, both dynamic and static
Plans
Alternative expansion designs
1 Electric supply conditions, calculation of daily and yearly load curves
2 Loads on system centres
3 Equivalent technical-economic characteristics of power generating installations
Models taking into account statistics of
a system needs for emergency reserve
b system needs for reserve in case of overhauls and running repairs
c typical day loads on the system
d off-balance loads on system centres
e design loads on main system networks
f technical-economic estimates of alternative system network schemes
g network cost estimates for specified station siting within the system
h calculation of network conditions, short-circuit currents and stability
i reliability analysis of system networks
j overall technical-economic indices of system operation
Models taking into account dynamics of
a power and energy balances of electric power systems and centres
b technical-economic estimates of alternatives for network expansion of electric power systems
c network cost estimates for specified alternative siting and construction priorities of power stations
d estimates of capital, running and discounted costs over the system expansion period under study
e overall technical-economic indices of system development

Fig. 1.12

With any problem-solving technique, it is important to set the limit of model precision desired. A matter of crucial importance to all system optimization is that a model need not be more precise than is actually necessary. This requirement in fact applies to the necessary and sufficient precision of input (design) data, model proper, solution technique and realization (the last point, although it is not related to modelling directly, has a decisive effect on it). Rather than cover the whole of the problem, we shall dwell on the precision of the model proper.

It is customary to state that a model should give a sufficiently complete and faithful representation of the basic system variables and functions. However, the words 'a sufficiently complete and faithful representation' are only a front to cover up a poor knowledge of the mathematical requirements any model ought to comply with.

At present, work has only been started to determine these requirements, and instead of a consistent system we have to be content with so-called general considerations in many cases. It is obvious that the success or failure of mathematical modelling will in the future be greatly dependent on how this problem is tackled.

Naturally, the precision of a model should be correlated with that of input data and of the realization planned. It is convenient to handle the precision of information separately from that of the model, because it often happens that some detail is deliberately removed from the system characterization in comparison with what the available information gives, so as to facilitate solution. An example is the use of linear models to solve nonlinear problems. In a wider sense, this is true of not only deterministic, but also of stochastic models. It is not always permissible to ignore detail, however. Whether or not this may be done should be decided according to the purpose of the analysis at hand. For example, whether or not a deterministic model may be used in a particular case can be determined by first using a probabilistic model (notably the random test method) for analysis. This analysis can outline a domain of optimal solutions, give a measure of errors that might arise due to the use of deterministic models, and settle the question as to the usability of deterministic models in particular cases.

The quality of analysis is directly dependent on how accurately we know the form and numerical values of distributions of random quantities. Unfortunately, the problem has not yet been investigated with sufficient detail. In fact, one of the weaknesses in optimal planning is the lack of techniques that would derive sufficiently precise information from analysis and yield relatively precise characteristics.

In a way, the reasonable precision of a model is related to that of realization. In this connection, it can be pointed out that a comparison of any design and its realization would immediately bring to light some regularity in the manner the plant actually built differs from what it looked like in the plans (of course, we omit subjective miscalculations and failure of judgement). This is what might be called lack of precision in realization.

It can take different forms: in terms of startup dates, in the quantity and kinds of materials used, the plant mix, parameters and characteristics, etc. It appears that this lack of precision in realization must be taken into account in model-building. Unfortunately, this approach is only in embryo.

Until a consistent system of requirements for model precision comes along, there is a tentative recommendation which, although based on fairly trivial considerations, may in many cases lead to an adequate model: a model may be taken as being adequate, if its error is by an order of magnitude smaller than that of input data.

1.4. Three Groups of Tasks in Control and Management of Power Systems

The tasks involved in the control and management of an interconnected electric power system are shown in Table 1.2. As is seen, the first group covers the direct operational control of the system, that is, what the control engineer has to do as a matter of his daily routine. The second group includes tasks which involve short-term scheduling, including the near future. The third group is that of long-term cycles, that is, long-term planning and forecasting. Unfortunately, practical terminology is in a state of confusion, and this is especially true of the delineation between planning and forecasting. There is an obvious shift of meaning when one speaks of planning with regard to the level of operation for a day, or of forecasting with regard to loads for the next month. Yet in most cases forecasting refers to longer and planning to shorter time periods.

The problems that arise in system control and management are solved on models, using suitable forms of programming. In general, a model is an intermediate arrangement (or process) which is used for an analysis of the prototype arrangement (or process). In economics, a model is usually a set of equations based on certain assumptions and describing the national economy as a whole or a part thereof. A model should concentrate on relevant features of the prototype, while removing from consideration anything which defies rigorous analysis. The point about relevant and irrelevant features usually causes a lot of argument. There is a body of opinion maintaining that the more a model contains the better. This can hardly be so, for an engineering analyst usually seeks to bring out main, or governing, features in the problem at hand, and not simply to cover as many factors as he can. In his dispute with the proponents of this "maximalist" view, Academician N. P. Fedorenko, an expert in mathematical economics, argues that the art of modelling is to grasp the essentials without overcrowding the model. Quoting T.J. Williams, he points out that in any situation a fool can list more decisive factors than God himself. For example, Newton ought to have noted the fact that the planets are somewhat flattened at the poles, that there are mountains on some of them, that at least one has trees, termites and other things, both pleasant and unpleasant. It takes common sense to break off a listing of these factors at the right moment.

Table 1.2

Three Groups of Tasks in Control and Management of an Interconnected Power System

Group I	Group II	Group III
Direct control (from seconds to hours) *Normal steady state* (1) optimal system regulation on hourly basis, including active and var dispatch at stations and substations equipped with reactive power sources (RPS); (2) frequency and power interchange control over tie lines (including limits of interchange); *Normal transient state* (3) routine system switching operations; (4) frequency and active power interchange control over tie lines (including interchange limiting); (5) optimum system scheduling on hourly basis; (6) optimum economic and var dispatch at stations and substations with RPS; (7) voltage control at system centres; (8) load limiting and shedding (scheduled outages) — order to lower control levels; *Abnormal state* (9) integrated operation of automatic failure clearing means, including: – switching in networks (system sectionalization and area isolation); – automatic throwover on intersystem and trunk lines; – emergency load shedding; – control of mechani-	*Short-term management cycle* (from days and weeks to months) (1) forecasts on active and reactive loads at system centres (on daily and monthly basis); (2) forecasts on river runoff; (3) calculation of static operating states of the system; (4) calculation of short-circuit currents; (5) calculation of protective relay settings (compiling a selectivity chart); (6) selection of gains for regulator (including excitation, speed, transformers, etc.); (7) selection of failure-preventive control settings; (8) analysis of static stability and determination of maximum power interchange over power transmission line under static stability; (9) transient analysis and determination of maximum power interchange over transmission lines under dynamic stability; (10) system reserve dispatch (monthly, weekly, daily), including: – selection of plant mix; – selection of plant for cold stand-by; – maintenance scheduling (on monthly basis); (11) determination of interoperation logic for automatic controls and regulators in case of emergency; (12) determination of logic for restorative conditions; (13) inflow, usage and	*Long-term management cycle* (from one to several years) (1) forecasts on active and reactive loads at system centres; (2) forecasts on river runoff; (3) forecasts on static load characteristics for system centres; (4) determination of incremental curves; (5) processing of emergency and failure statistics (forecasts on sudden failures); (6) calculation of static operating states for the system; (7) calculation of short-circuit currents; (8) calculation of protective relay settings (compilation of selectivity chart); (9) selection of gains for regulators (excitation, speed, transformers, etc.); (10) selection of settings for automatic failure-clearing means; (11) static stability analysis and determination of maximum power flow over transmission line under static stability conditions; (12) transient analysis and determination of maximum power flow under dynamic stability conditions; (13) reserve dispatch (on yearly basis), including: – selection of plant mix in operation; – selection of plant for cold stand-by; – main plant scheduling;

Table 1.2 *(Continued)*

Group I	Group II	Group III
cal and electrical braking for generators; – emergency regulation of generator excitation and voltage at system centres; – active power limiting at stations; (10) regulation of frequency and active power flow over transmission line in emergency; *Post-fault or restorative (transient and steady) state* (11) integrated operation of automatic state-restorative means, including: – switching in network, resynchronization of isolated system parts, restoration of disrupted connections (including loads); – throw-over to hot and cold stand-by sources; (12) frequency and active power flow control; (13) system dispatch, including: – economic and var dispatch for stations and substations with RPS; – voltage regulation at system centres	spillage scheduling for hydro reservoirs and cascaded stations (on weekly and monthly basis); (14) calculation of optimum system operation over cycle; (15) outage scheduling; (16) monthly power balance and power generation for system areas	(14) determination of interoperation logic for automatic controls and regulators in emergency; (15) determination of logic for post-fault (restorative) system state; (16) calculation of automatic frequency regulation (on semi-annual basis); (17) calculation of overvoltages (internal and atmospheric); (18) inflow, usage and spillage scheduling for hydro reservoirs and cascaded stations (on annual and multi-annual basis); (19) calculation of maximum and optimum levels of operation; (20) scheduled load outage; (21) annual power balance and power generation for system areas

In practice, a listing should be discontinued as soon as common sense prompts that the model covers enough features for the experimenter to learn anything about the problem he is interested in.

In our case, the object of analysis is an economic entity, such as an inter-sector flow structure. If a model is built to reflect the essential features of this structure, we may expect that the analysis will be correct. And again it must be stressed that a model should concentrate on features which are decisive in the problem at hand. Model-building must be creative and sufficiently flexible so as to fit the model to whatever can be decisive in a particular case and for a particular form of analysis.

Among the problems facing the control and management of large power systems, the most important is the search for ever better methods of power-system optimization. It is essential to work out a correlated complex of

mathematical models that would reflect the actual hierarchical structure of large systems. This is an urgent problem, the more so that we are still lacking reliable methods for building such models.

The second important problem is to devise methods for the acquisition and processing of data essential to power-system expansion. Here, it is important to set up a consistent system which would gather, process and present data, evaluate the effects of data corruption, and advance both theoretical and experimental work on data transmission. These are all vital aspects of a mammoth problem, for if a control system is to be closed-loop and to work efficiently, in addition to computers it must have data to be put in the computers and uncorrupted data at that. These data must be collected, processed, and transmitted, which in turn necessitates the use of appropriate data links and techniques for data acquisition, reduction and algorithmization that would make them acceptable to computers. To quote Academician V. M. Glushkov, a present-day computer is like a bottle with a large body but a narrow neck; you can store much in it, but it takes you some effort to put anything into or out of it. Therefore, the bottleneck must be made wide for free exchange of data. A concommitant problem is how to use data derived from information known to be incomplete in advance or departing from reality in some way. In other words, there is a problem of how to make reliable data out of incomplete and unreliable input.

Work on the theory and practice of power-system control and expansion needs a material basis, above all computer centres, data processing facilities and all that will make an automatic management information system, or an MIS for short. In the Soviet Union, such management information systems are being developed on a large scale, so as to cater for all industries, including the power industry.

The problems that need solution before a reliable MIS can be set up for the power industry are interrelationships between large power systems and the national economy and the biosphere, that is, the *environment* in the broad sense of the word.

Finally, we need criteria of optimality which would reflect the dynamics of power-system expansion and take proper account of advances in fundamental and applied research. No accepted criteria of the kind exist yet, and their formulation and refinement is a vital, although evasive problem.

We have to leave out of our discussion one more important problem — that of the methods and procedures to be used in closing-cost finding and the price structure as related to the power industry, although it has a certain bearing on power-system control and management.

1.5. Planning and Forecasting in the Power Industry at Present

At present, planning for the power industry is mainly based on technical and economic surveys and expansion schemes for the industry as a whole (in the form of technical and economic reports) and for the various indi-

vidual power systems (from the USSR electric power supergrid down to local supply networks). In both cases, the procedure is as follows.

Stage I. A technical and economic report is drawn up for a period of 15 to 20 years. In this report:

(1) principal directions are mapped out for the expansion of major power systems (such as the USSR and European systems);

(2) levels of power demand are defined for the country as a whole and for major regions, on the basis of tentative estimates of economic growth supplied by economic planning bodies;

(3) estimates are given for operating states, the generating capacities to be activated, and the loads to be supplied (peak or base);

(4) potential energy resources are revealed in the various regions and the country as a whole, and the level of performance is defined for different stations (hydro, thermal, nuclear, and heat-generating);

(5) from a comparison of alternative schemes for their technical and economic features or from tests of appropriate mathematical models, an optimum generating plant mix is defined, that is, an optimum proportion of contributions from heat-and-power stations, hydro stations and thermal plants burning different fuels. Consideration is given to the comparative effectiveness of carriage for different forms of fuel and electricity;

(6) the level of operation is fixed for the stations with the plant mix defined above, and on its basis the plant mix and station siting for the individual regions are refined; an estimate is formed for the level of operation of inter-system connections;

(7) on the basis of the load breakdown among power centres, balance sheets are compiled and major directions are defined for the expansion of the electric networks within the integrated electric power system and the pan-energy system. A configuration is decided upon for the principal networks. Due consideration is given to the joint operation of major power interconnections.

The technical and economic report provides a basis for recommendations to be used in the siting of thermal power stations and substations.

Stage II. An expansion scheme is drawn up for integrated and interconnected power systems for a period of 5 to 10 years ahead. As part of such a scheme:

(1) refined and detailed data are given on the levels of demand for electricity and heat;

(2) final figures are stated for the capacities of heat-and-power stations and thermal plants, their siting, and due dates of their construction in the light of the actual capabilities of contracting organizations;

(3) connection schemes are chosen for the electric networks, their levels of operation are defined, and requirements are formulated for their automatic fault-clearing and control facilities;

(4) reactive power sources are sited and conditions for voltage control are defined;

(5) requirements are formulated for switching apparatus. After connection schemes have been developed for interconnected systems, a similar job is done for individual power systems where emphasis is mainly placed on the expansion of the individual distribution networks;

(6) an analysis is carried out into the growth of power demand and the merger of power systems and stations, and practical problems arising in the course of design work are resolved.

The practical needs of planning and forecasting necessitate further work on the following aspects:

(1) a classification of stages in power-system planning and formulation of the tasks to be achieved at each stage;

(2) techniques for determining demand at major centres of large interconnections on the basis of outline estimates (from a forecast on the overall growth in electricity demand for an interconnection as a whole and from incomplete information on consumer mix);

(3) techniques for evaluation of anticipated electricity and heat demand, that would take account of changes in consumer mix and consumption pattern and utilize probabilistic concepts in view of inaccuracy of forecasts;

(4) techniques for an outline evaluation of the expansion of heat and power capacities at the nodes of large interconnections;

(5) an improved theory of power-system structure optimization. This theory should take into account the dynamics of system expansion and lay groundwork for techniques to determine an optimum breakdown of contributions from various energy sources and types of power plant, recalling that input data are partly probabilistic in character. In the light of the foregoing, it appears imperative to develop:

(a) refined general mathematical models so as to determine optimum directions for power-system development;

(b) a system of partial mathematical models to facilitate comparison of alternative plans for power-system expansion and to select optimal characteristics for the individual elements of power systems and plants (for example, optimum unit generating capacity, optimum network layout, etc.);

(c) techniques that would enable the system structure to be chosen in accordance with expected conditions of system operation;

(6) the necessary corrections for the partly probabilistic character of input data at the time of network design;

(7) criteria of an optimum configuration and parameters for a complex electric network;

(8) techniques to determine the optimum power handling capacity of tie lines within complex power systems;

(9) methods to analyse parallel operation of large power systems incorporating electric networks of complex configuration;

(10) methods for evaluating the system reliability including those which define:

(a) the required optimum reserve of power and energy within each system;

(b) reliability criteria for complex interconnections;
(c) reliability criteria for electric networks;
(11) refined evaluation of the effects that changes in technical and economic characteristics may have on the development of power plants and on the length of the time periods to be covered at each particular stage of power-system planning.

There are also more general tasks, some of which have already been mentioned, namely:

(1) study into the principal directions in which power systems in the Soviet Union may develop up to the year 2000; and evaluation of the prospects of using novel types of power generating and converting means in the systems;

(2) analysis of the structural changes that are likely to take place in the power consumption pattern in view of the anticipated economic growth; evaluation of the level of power demand for the country as a whole and for the major regions at different times within the period under consideration;

(3) determination of rational limits for the use of different fuels for power generation and the 'closing' indicators of fuel supply in the various parts of the country;

(4) formulation of consistent technical and economic indicators for electric stations, fuel sources and fuel carriers and techniques for their determination for the purpose of planning and forecasting;

(5) analysis of trends in the development of power generating plant and optimal limits thereof (capacity of turbine and boiler units, flexibility of operation, etc.);

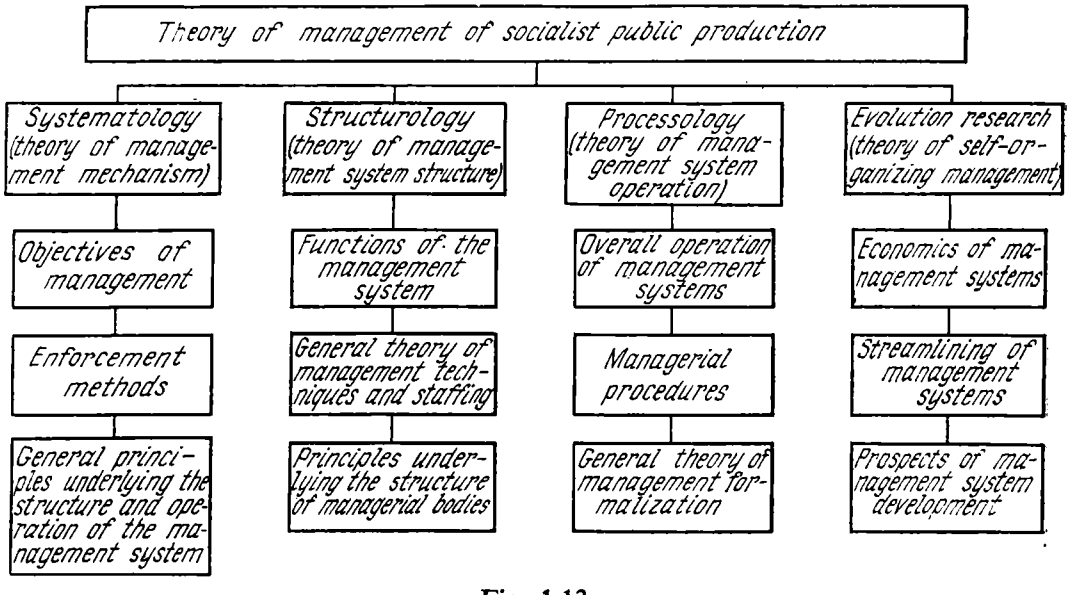

Theory of management of socialist public production
Systematology (theory of management mechanism)
Structurology (theory of management system structure)
Processology (theory of management system operation)
Evolution research (theory of self-organizing management)
Objectives of management
Functions of the management system
Overall operation of management systems
Economics of management systems
Enforcement methods
General theory of management techniques and staffing
Managerial procedures
Streamlining of management systems
General principles underlying the structure and operation of the management system
Principles underlying the structure of managerial bodies
General theory of management formalization
Prospects of management system development

Fig. 1.13

(6) formulation of conditions for the joint operation of the power systems that make up the country's integrated power supergrid, determination of power flows, selection of the highest voltage level and of main network configurations.

The various aspects of management and control theory related to the forecasting, planning and operation of power systems and to the power industry as a whole, which have been discussed in this chapter are only a part of the more general theory of socialist public production*. An idea about the other parts of this theory can be gleaned from the chart of Fig. 1.13.

BIBLIOGRAPHY TO CHAPTER 1**

1. Melentyev L. A., editor. *Mathematical Modelling in the Power Industry* (Metody matematicheskogo modelirovaniya v energetike). Complied by the East-Siberian Publishing House. Irkutsk, 1960.
2. Melentyev L. A. *Optimization and Control of Large Power Systems* (Optimizatsiya i upravlenie v bolshikh sistemakh energetiki). East-Siberian Publishing House. Irkutsk, 1970.
3. Venikov V. A. *Use of Cybernetics in Electric Power Systems* (Primenenie kibernetiki v elektricheskikh sistemakh). Collected volume "Cybernetics in the service of communism" (Kibernetika na sluzhbe kommunizma), vol. 1, GOSENERGOIZDAT, 1961.
4. Venikov V. A. *On the Objectives of the Theoretical Work Related to the Problem of an Automatic Control System for the USSR Integrated Electric Power Supergrid* (O zadachakh nauchno-teoreticheskikh razrabotok svyazannykh s problemoy ASU EEES SSSR). Proceedings of the USSR Academy of Sciences. "Energetika i transport", 1972, No. 2.
5. Dobrov R. M. *Forecasting in Science and Technology.* (Prognozirovanie nauki i tekhniki). "Nauka", 1969.
6. Lisichkin V. A. *Scientific Forecasts for Individual Industries* (Otraslevoe nauchnoe prognozirovanie). "Ekonomika", 1971.
7. Afanasyev V. R. *On the Intensification of Socialist Society (Problems of Cooperation between Science, Technology and Management)* (Ob intensifikatsii sotsialisticheskogo obshchestva (problemy vzaimodeystviya nauki, tekhniki i upravleniya), "Mysl", 1969.
8. Popov G. Kh. *Methodological Problems of the Theory of Management of Socialist Public Production* (Metodologicheskie problemy teorii upravleniya sotsialisticheskim obshchestvennym proizvodstvom). Collected volume "Sum-totals of science and technology" (Itogi nauki i tekhniki). Management organization. Vol. 2, 1973.
9. Pavlenko A. S. et al. *The Power Industry of the Soviet Union in 1971-1975* (Energetika SSSR v 1971-1975), "Energiya", 1972.
10. Gomoshinsky V. G. *Forecasting Practice* (Praktika prognozirovaniya). "Znanie", 1972.

* For further detail, see ref. [8].

** This is a list of recommended reading, and not that of references used in writing Chapter 1.

CHAPTER 2

OPTIMIZATION OF ELECTRIC POWER SYSTEMS DURING EXPANSION PLANNING

2.1. Optimization of Electric Power System Expansion as a Cybernetic Problem

As already noted, electric power systems are large cybernetic systems which are expanding continually. They are cybernetic above all because their expansion is purposeful and optimal both economically and technically. Failure of power generation to meet power demand may have detrimental effects on the national economy, as it would entail a slow-down in industrial growth, population welfare, and labour productivity. The fact that electricity is extremely versatile, coupled with the ever wider use of electric power in industry, transport, agriculture and for domestic needs, makes the timely expansion of the power industry a matter of paramount importance to all of the national economy. Because the nation puts enormous sums of money and material resources into the power industry, it is vital that the power systems and their interconnections should develop optimally from an economic point of view.

The optimization of expansion plans for an electric power system is a very complex problem involving the use of mathematical programming and digital computers on a large scale. Typically, it consists of the following stages:

(a) input (design) data are collected;

(b) the objective function is formulated;

(c) an algorithm and a program are compiled for a digital computer to minimize the objective function.

The optimization of system expansion is a dynamic problem in the sense that it cannot be solved for any specific instant of time (say, a year). Capital outlays must be optimized for a sufficiently long time period ahead (10 to 20 years), because the optimal decisions for the individual years of the period may be conflicting. For example, a power installation which looks optimal, if started during the early years, may well end up as a liability if its start-up is deferred until the end of the same period, and vice versa.

As regards the expansion of an electric power system, the solution of an optimization problem needs the following design or input data:

(a) on dynamics of power demand at the various nodes of the system for each year of the period under consideration, including maximum capacity, energy consumption, total load structure and duration;

(b) on the energy resources that can be utilized by the system directly

or through transfer from other systems, including limits of utilization for these resources, the economical indicators of the resources and of their transfer (over power transmission lines, by pipelines, by road, etc.);

(c) on the economic and technical characteristics of power generating units and network elements, including those already in operation;

(d) on the reliability of the power system, including fault statistics on system elements, usage and spillage statistics in the case of hydro stations, the cost of loss due to interruptions in electricity supply, etc.;

(e) on the form and character of the objective function.

As is seen, many of the input data can only be obtained from forecasts whose accuracy must of necessity fall off rapidly as the forecaster looks farther into the future. The expenditures associated with the later years of the period under consideration decrease too due to their payoff discount (to be explained later).

Consider the input data by groups.

Data on electric power demand for individual centres and for the system as a whole are derived from forecasts of demand by various groups of consumers, and from statistics relating annual demand to other demand indicators (maximum, average and minimum loads, typical day's load curves).

Data on the annual growth of power demand for individual industries and the national economy as a whole are based on their respective development plans over the period under consideration, notably on the planned production expansion and planned reduction in power consumption per unit of output.

Data on the annual increase in power demand for domestic uses, public utilities, entertainment and the like are based not only on their respective development plans, but also on statistics of the natural growth of power consumption and population.

Data on energy resources are usually based on the country's optimized fuel and energy balance which takes account of the fuel needed not only for power generation, but also for other uses. Concurrently, an optimal scheme is evaluated for the transfer of power both in terms of the geography of the points of dispatch and reception and in terms of the energy transferred, and the fuels themselves are assessed economically.

Data on the economic and technical characteristics of power generating plant and network elements are usually derived from their designers' and makers' data sheets.

Failure rates are obtained from statistical records. In this connection, it is important to remember that failure rates are usually high during the shake-down period and tend to level off or even decrease during the useful life of the same equipment.

Data on the cost of loss due to interruptions in electricity supply are usually specific to a particular system. The figure adopted by Soviet designers for reserve capacity optimization is 0.6 rouble/kWh, but there are indications that it will decrease with time.

In expansion optimization, the objective function is the sum of all discounted costs of the power installations to be built during the period under consideration and also the discounted costs of existing units, over the period of existence. These discounted costs include capital or investment costs, maintenance and operating costs.

The objective function may then be written as

$$Z_0=\sum_{t=0}^{T-1} I_t(1+p)^{\tau-t}+\sum_{t=0}^{T_s-1} O_t(1+p)^{\tau-t-1} *$$

where I_t = investment costs of all installations existing or built in year t
O_t = operating costs of all installations existing or built in year t
T = optimization period
T_s = service life of all system elements (of course different for different elements, but assumed to be the same for all and equal to infinity)
p = payoff or discount factor equal to 0.10 to 0.12, with which all expenditures are discounted in terms of a reference year, τ

After some manipulations,

$$Z_0=\sum_{t=0}^{T-1} [I_t(1+p)^{\tau-t}+O_t(1+p)^{\tau-t-1}]+O_{T-1}\sum_{t=T}^{\infty}(1+p)^{\tau-t-1}$$

where the operating costs of all power installations built by the end of the period under consideration are assumed to be constant and equal to those during the last year, O_{T-1}.

The second term of the expression may be written as

$$O_{T-1}\sum_{t=T}^{\infty}(1+p)^{\tau-T-1}=O_{T-1}[(1+p)^{\tau-T-1}+(1+p)^{\tau-T-2}+\ldots]$$

$$=O_{T-1}(1+p)^{\tau-T}\left[\frac{1}{1+p}+\frac{1}{(1+p)^2}+\ldots\right]=O_{T-1}(1+p)^{\tau-T}(1/p)$$

where $1/p$ is the sum of the terms of the geometric progression in the brackets.

Hence,

$$Z_0=\sum_{t=0}^{T-1} [I_t(1+p)^{\tau-t}+O_t(1+p)^{\tau-t-1}]+O_{T-1}(1+p)^{\tau-T}(1/p)$$

If, instead of Z_0, we minimize $Z_0p=z$, the new objective function will be

$$z=p\sum_{t=0}^{T-1} [I_t(1+p)^{\tau-t}+O_t(1+p)^{\tau-t-1}]+O_{T-1}(1+p)^{\tau-T}$$

* In this function, it is assumed that capital is invested during the last year of construction, that is, before start-up, and the annual maintenance and operating charges of an installation remain a constant percentage of capital cost.

Since the discount period, τ, is chosen arbitrarily, it may well be set equal to zero, and then

$$z=p\sum_{t=0}^{T-1}[I_t(1+p)^{-t}+O_t(1+p)^{-t-1}]+O_{T-1}(1+p)^{-T}$$

The above expression gives the sum of discounted costs for an expansion optimization problem, under the assumptions made.

In the case of a purely static problem, when $t=0$ and $T=1$, we may write

$$z=pk_0+pO_0(1+p)^{-1}+O_0(1+p)^{-1}=pk_0+O_0$$

In the above form, the objective function is used in static optimization problems.

Capital investments are a function of station capacity and may be approximated by a nonlinear relation of the form

$$A+BP$$

where A is the fixed component independent of installed capacity P, and B is a constant coefficient. The value of A is especially high for hydro stations. This is why, the first generating sets of a hydro station involve higher costs than the succeeding ones; as a rule, the latter are less expensive than the comparable sets of thermal power stations.

The capital costs of the main network elements depend on their capacity or throughput in a similar way.

Fuel costs for thermal power stations depend in a complex nonlinear fashion on installed capacity, time of utilization, and system structure. As will be recalled, the cost of transmission losses is a quadratic function of the power transmitted. Also, the loss of power depends on the level of operation and operating state at the time under consideration. As a rule, the energy loss problem is simplified by introducing an equivalence factor, τ, which defines the number of hours during which the maximum power loss would give actual energy losses that would be a quadratic function of the maximum power transmitted.

For power transmission lines under construction, the loss of power and energy may be arranged to depend linearly on maximum flow by the choice of an appropriate current density. As will be recalled, the power loss is I^2r. After simple manipulations at $\cos\varphi=1$, we have

$$I^2r=I^2\varrho l/q=I\varrho li_e=i_e(P/V)\varrho l$$

where ϱ = resistivity of the conductors
l = line length
q = cross-sectional area of the conductors
i_e = economic current density equal to I/q
P = transmitted power
V = operating voltage

Renovation and maintenance costs are usually assumed to be in proportion to the investment cost.

2.2. Expansion Optimization from a Mathematical Point of View

As is seen from Sec. 2.1, the objective function in an expansion optimization problem is a complex nonlinear function of a very great number of unknowns some of which are integers (such as the number of generating units at a station, power-line voltage levels, cross-sectional area of conductors, etc.). More specifically, the unknowns of an expansion optimization problem are:

(1) the optimum commissioning date for a new power installation (a station, a transmission line);

(2) the optimum type of installation to be set up (capacity of generating units, kind of fuel, steam pressure and temperature, voltage, conductor cross section, number of circuits in the line, etc.);

(3) the number of units (circuits) to be put in operation every year over the period under consideration.

In the long run, the solution should define the optimum structure of the stations and the network configuration for each year of the period considered. At present, this type of problem, when formulated rigorously, cannot be solved, even by the latest makes of electronic computers. As a way out, a number of simplifying assumptions have to be made so as to build mathematical models yielding sufficiently accurate solutions.

Mathematical models and their blocks use different techniques of mathematical programming. *Mathematical programming* is a division of computational mathematics which handles various classes of optimization problems by supplying appropriate algorithms, investigating conditions for convergence, and testing a solution for uniqueness.

We shall dwell in sufficient detail on only some techniques of mathematical programming applicable to the practical problems of power-system expansion optimization.

All of these techniques may be classed into linear and nonlinear.

Linear programming techniques apply when the objective function depends linearly on the sought unknowns and all constraint equalities and inequalities written in terms of the unknowns are linear. This limits the area of application for linear programming techniques. However, they are so efficient that we may use them in some nonlinear optimization problems, as well. Linear optimization problems yield a unique solution.

Nonlinear programming techniques apply when the objective function is nonlinear, and so are the constraint equalities and inequalities. Nonlinear optimization problems may have several solutions; that is, they are multiply extremal. Because of this, in using nonlinear programming techniques, it is important to determine a global extremum (the least minimum or the highest maximum), which is not always feasible. Experience with nonlinear programming is still scanty, and there are limitations on its use.

According to the form of unknowns, both classes of mathematical programming may be further divided into sub-classes of continuous, discrete and integer programming. The last two subclasses may be combined, be-

cause problems involving integer programming may be transformed to those handled on the basis of discrete programming.

This chapter will examine the following techniques of linear programming: the simplex algorithm method, the simplex method, the modified simplex method, Gomory's integer linear programming method, and a classical method known as the transportation problem. Of all techniques of nonlinear programming, our concern will be with the gradient method, the dynamic programming method, and the branch and bound method.

As already noted, mathematical models can be of the *system-wide (global)* type or the *partial (block)* type. With a global model, the objective function is minimized as a whole. With a block model, each block performs a limited task of optimization or estimation, and the objective function is minimized as a whole through interaction of the individual blocks on the basis of iteration.

Two forms of mathematical models can be used in expansion optimization, namely *optimization* models and *estimation* models. The former are intended for a direct formulation of an optimum power-system expansion plan. This plan should fix the commissioning date (year) for each installation, the characteristics of the generating plant installed (steam pressure and temperature, generating-unit capacity, line voltage, cross-sectional area and number of conductors in a line, etc.), and the capacity of each installation to be put in service every year within the period considered. The direct formulation of an optimum expansion plan is complicated by the fact that it involves a great number of unknowns which are usually constrained to take discrete or integer values (commissioning year, plant parameters), and also by the fact that the objective function is nonlinear. This is why the optimization models in current use are built as linear models and handled by some modifications of linear programming. The inaccuracies inherent in this approach limit the area of application for optimization models. They are mainly resorted to in a preliminary survey of the effect that various factors may have on the optimality of a particular expansion scheme and in order to estimate the effectiveness of individual units put in service. In this way, planners are in a position to narrow down the area of search to fewer alternatives that may potentially be optimal.

In contrast, estimation models are nonlinear, that is, they reflect the fact that both the objective function and the constraints are nonlinear. Models in this class evaluate economically a considerable number of alternatives that the designer may submit for consideration. As a rule, expansion alternatives are concerned solely with power stations and leave out the costs of fuel and networks and reliability. It is the goal of an estimation model to give an estimate of the discounted costs of fuel and networks and the cost of loss due to interruption in electricity supply. When added to the construction cost of the stations, these costs yield an overall economic estimate of a particular alternative plan.

The number of alternatives eventually submitted for consideration is limited above all by the designer himself who discards all plans that are

obviously uneconomical. The number of alternatives can be further reduced if expansion schemes are already included in the construction plans for new stations. Then alternatives will mainly be those of priority in the construction of particular stations.

The discounted construction and operating costs (leaving out fuel costs) usually account for 40 to 50% of the total cost. This cost item may be specified by the designer for each alternative design or computed by an appropriate block in the model.

The fuel costs account for 35 to 45% of the total cost. Thus, taken together, the construction and fuel costs are 75 to 90% of the total. The construction and operating costs for transit networks (ignoring the loss in the networks) are 5 to 10%. The cost of loss to consumers due to interruption of electricity supply is 2 to 4%.

An estimation model usually comprises a network block, a system-state block, and a loss-of-supply block, each handling the respective cost item. In addition, these model blocks may be used to handle partial design tasks independently.

2.3. Simplex Algorithm

The linear programming techniques used in linear mathematical models (see Sec. 2.2) are mainly modifications of the simplex method. We shall consider the simplex algorithm which is the basis of the simplex method.

It is applicable to the optimization of a linear objective function of many variables related by linear constraint equalities on the condition that the constraint equations and the objective function are in their canonical form (to be defined later).

Let the objective function be a linear function of n unknowns $x_1, x_2, \ldots, x_n$:

$$z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \tag{2.1}$$

where

$$x_i \geqslant 0, \quad i = 1, 2, \ldots, n \tag{2.2}$$

The constraint equations have the form

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}\right\} \tag{2.3}$$

where $m < n$. At $m = n$, the variables are uniquely defined by simultaneous equations (2.3), and z is defined, likewise uniquely, by Eq. (2.1). All variables are non-negative on the strength of Eq. (2.2). The unknowns which have a coefficient of unity in one of Eqs. (2.3) and do not appear in any other equations, Eq. (2.3), or in the objective function, Eq. (2.1), will be called *basic variables*. The remainder of the variables will be called *non-basic*.

If there are m basic variables, that is, as many as there are constraint equations, then Eqs. (2.3) and the expression for z will take the following canonical form

$$\left.\begin{aligned} x_1 \quad &+a_{1,m+1}x_{m+1}+a_{1,m+2}x_{m+2}+\ldots+a_{1n}x_n &&=b_1\\ x_2 \quad &+a_{2,m+1}x_{m+1}+a_{2,m+2}x_{m+2}+\ldots+a_{2n}x_n &&=b_2\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ x_m &+a_{m,m+1}x_{m+1}+a_{m,m+2}x_{m+2}+\ldots+a_{mn}x_n &&=b_m\\ &c_{m+1}x_{m+1}+c_{m+2}x_{m+2} \quad +\ldots+c_nx_n-z=-z_0 \end{aligned}\right\} \tag{2.4}$$

where $x_1, x_2, \ldots, x_m$ are the basic variables.

If all non-basic variables ($x_{m+1}, x_{m+2}, \ldots, x_n$) are zero in value, then all basic variables ($x_1, x_2, \ldots, x_n$) will be equal to the free terms $b_1, b_2, \ldots, b_m$, and the objective function will be z_0. A solution in which all non-basic variables are zero in value will be called a *basic solution.*

A basic solution will be called feasible, if all basic variables are non-negative. A feasible basic solution satisfies Eqs. (2.2) and (2.4). The manner in which the systems of equations, Eqs. (2.1) and (2.3), can be re-cast into their canonical form will be examined in Sec. 2.4, in connection with the simplex method.

If the constraint equations and the objective function are in their canonical form, Eqs. (2.4), the function z can be minimized by the simplex algorithm as follows.

As is perfectly obvious, the minimality criterion for z requires that all coefficients c_j in Eqs. (2.4) be non-negative. If all $c_j \geqslant 0$, then any change in the non-basic variables, because they are non-negative, will cause z either to increase or to remain unchanged, which is a proof that the solution is optimal. If at least one of the coefficients in the system, Eqs. (2.4), is less than zero, $c_j < 0$, z can be minimized by maximizing x_j. Then the unknown x_j ceases to be non-basic and must be included with the basic variables, or introduced into the basis.

If several coefficients c_j are negative, the optimization procedure can be speeded up by choosing the minimum coefficient, that is, one having the highest absolute value. Let this minimum coefficient be c_s:

$$c_s = \min (c_j) \tag{2.5}$$

such that

$$c_s < 0$$

Also, the unknown x_s is introduced into the basis. The index s, that is, the unknown added to the basis, is chosen in the simplex algorithm from the condition (2.5). The introduction of x_s into the basis will cause z to decrease, because the product $c_s x_s < 0$.

It is important to determine the limit to which x_s may be increased. This limitation on x_s stems solely from the condition, Eq. (2.2), that all unknowns be non-negative. Let us find how the basic variables, $x_1, x_2, \ldots, x_m$, will

change due to the introduction of x_s with the remaining non-basic variables equated to zero. From Eq. (2.4) it follows that

$$\left.\begin{aligned} x_1 &= b_1 - a_{1s}x_s \\ x_2 &= b_2 - a_{2s}x_s \\ &\cdots\cdots\cdots\cdots \\ x_m &= b_m - a_{ms}x_s \end{aligned}\right\} \tag{2.6}$$

From inspection of Eqs. (2.6), the following two conclusions may be made:

(1) if all a_{is} are negative or equal to zero, that is, $a_{is} \leq 0$, x_s may be increased without bound and, as a consequence, the function z is not bounded below;

(2) if some a_{is} are positive, that is, $a_{is} > 0$, there are limitations on an increase in x_s. In such a case, the respective basic variables x_i will decrease and, as soon as at least one of them reaches zero, any further increase in x_s would be *illegitimate*. The first to reach value zero will be the basic variable x_i for which the ratio b_i/a_{is}, that is, the ratio of the free term to the coefficient in the s-column is a minimum, because x_s cannot be greater than b_i/a_{is} owing to Eqs. (2.2) and (2.6). Therefore, the limiting value of x_s is defined as

$$x_s = \min (b_i/a_{is}) \tag{2.7}$$

where $a_{is} > 0$.

With this value of x_s, one of the basic variables x_i takes on value zero, that is, the basic variable x_i is eliminated (from the basis). The index r of the basic unknown x_r eliminated from the basis is thus chosen from the condition

$$b_r/a_{rs} = \min (b_i/a_{is}) \tag{2.8}$$

where $a_{is} > 0$.

Thus, every step in the simplex algorithm involves:

(a) the selection of the non-basic variable x_s to be introduced into the basis, subject to the condition, Eq. (2.5);

(b) the selection of the basic variable x_r to be eliminated from the basis, subject to the condition, Eq. (2.8).

As a result, a new feasible basic solution is obtained, which gives the value of z diminished by the amount

$$|c_s x_s| = |c_s b_r/a_{rs}|$$

that is, a new diminished value

$$z_0^* = z_0 + c_s x_s < z_0$$

because $c_s < 0$ and $x_s > 0$.

Before proceeding with the optimization any further, the system of constraint equations and the objective function should again be reduced to their canonical form. To do this, the new basic variable x_s must be elimi-

nated from all equations, Eqs. (2.4), except in the r-row. In the last equation, the coefficient of x_s should be unity. This transformation is carried out as follows:

(a) the r-row of the system, Eqs. (2.4), that is the equation

$$x_r+a_{r,\,m+1}x_{m+1}+a_{r,\,m+2}x_{m+2}+\ldots+a_{rs}x_s+\ldots+a_{rn}x_n=b_r$$

is divided into the coefficient a_{rs}, called the leading term of the transformation, to give the value of x_s:

$$x_s=\frac{b_r}{a_{rs}}-\frac{x_r}{a_{rs}}-\frac{a_{r,\,m+1}x_{m+1}}{a_{rs}}-\frac{a_{r,\,m+2}x_{m+2}}{a_{rs}}-\ldots-\frac{a_{rn}x_n}{a_{rs}} \tag{2.9}$$

(b) Eq. (2.9) is substituted for x_s in all the other equations of the system, Eqs. (2.4), to reduce the system to a new canonical form:

$$\left.\begin{aligned}
x_1 \quad & +a_{1,\,m+1}x_{m+1}+\ldots+a_{1r}x_r+\ldots+a_{1n}x_n=b_1\\
x_2 \quad & +a_{2,\,m+1}x_{m+1}+\ldots+a_{2r}x_r+\ldots+a_{2n}x_n=b_2\\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
x_s \quad & +a_{s,\,m+1}x_{m+1}+\ldots+a_{sr}x_r+\ldots+a_{sn}x_n=b_s\\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
& x_m+a_{m,\,m+1}x_{m+1}+\ldots+a_{mr}x_r+\ldots+a_{mn}x_n=b_m\\
& c_{m+1}x_{m+1}+\ldots+c_rx_r+\ldots+c_nx_n-z=-z_0
\end{aligned}\right\} \tag{2.10}$$

where the x_s's are introduced into, and the x_r's are eliminated from, the basis. In manual computation, the value of x_s found from Eq. (2.9) is substituted in Eqs. (2.4), and the like terms are cancelled.

On a computer, the coefficients and free terms of Eqs. (2.4) are computed as follows:

(a) Eq. (2.9) is substituted for x_s in the ith row to yield

$$x_i+a_{i,\,m+1}x_{m+1}+\ldots+a_{is}\left(\frac{b_r}{a_{rs}}-\frac{x_r}{a_{rs}}-\frac{a_{r,\,m+1}x_{m+1}}{a_{rs}}-\ldots-\frac{a_{rn}x_n}{a_{rs}}\right)+\ldots+a_{in}x_n=b_i \tag{2.11}$$

After the similar terms are cancelled, the ith row, Eq. (2.11) takes the form

$$x_i+\left(a_{i,\,m+1}-\frac{a_{is}a_{r,\,m+1}}{a_{rs}}\right)x_{m+1}+\ldots+\left(a_{ij}-\frac{a_{is}a_{rj}}{a_{rs}}\right)x_j-\frac{a_{is}}{a_{rs}}x_r+\ldots+\left(a_{in}-\frac{a_{is}a_{rn}}{a_{rs}}\right)x_n=b_i-b_r\frac{a_{is}}{a_{rs}} \tag{2.12}$$

(b) Eq. (2.9) is substituted for x_s in the expression for z, to yield

$$c_{m+1}x_{m+1}+\ldots+c_s\left(\frac{b_r}{a_{rs}}-\frac{x_r}{a_{rs}}-\frac{a_{r,\,m+1}x_{m+1}}{a_{rs}}-\ldots-\frac{a_{rn}x_n}{a_{rs}}\right)+\ldots+c_jx_j+\ldots+c_nx_n-z=-z_0$$

or

$$\left(c_{m+1}-c_s\frac{a_{r,\,m+1}}{a_{rs}}\right)x_{m+1}+\ldots-c_s\frac{x_r}{a_{rs}}$$
$$+\ldots+\left(c_j-c_s\frac{a_{js}}{a_{rs}}\right)x_j+\ldots+\left(c_n-c_s\frac{a_{rn}}{a_{rs}}\right)x_n-z=-z_0-c_s\frac{b_r}{a_{rs}} \qquad (2.13)$$

The following general equations can be used to transform the coefficients a_{ij} (where i is the row containing the basic variable of interest, j is the column containing the respective non-basic variable) and all free terms b_i:

for all rows, except s, and all columns, except r

$$\begin{aligned} a_{ij}^*&=a_{ij}-a_{is}a_{rj}/a_{rs}\\ b_i^*&=b_i-b_ra_{is}/a_{rs}\end{aligned} \qquad (2.14)$$

for the s-row, as follows from Eq. (2.9)

$$\begin{aligned} a_{sj}^*&=a_{rj}/a_{rs}\\ b_s^*&=b_r/a_{rs}\end{aligned} \qquad (2.15)$$

for the r-column, as is seen from Eqs. (2.9) and (2.12)

$$\begin{aligned} a_{ir}^*&=-a_{is}/a_{rs}\\ a_{sr}^*&=1/a_{rs}\end{aligned} \qquad (2.16)$$

where the asterisk (*) designates a new value of the coefficient or free term.

The following equations can be used to transform the c_j's and z_0:

for all columns except for r-column, see Eq. (2.13)

$$c_j^*=c_j-c_sa_{rj}/a_{rs}$$

for the r-column, see Eq. (2.13)

$$\begin{aligned} c_r^*&=-c_s/a_{rs}\\ z_0^*&=z_0+c_sb_r/a_{rs}\end{aligned} \qquad (2.17)$$

Equations (2.14) through (2.17) are used to re-cast Eqs. (2.4) into a new canonical form after x_s is introduced into, and r_x is eliminated from, the basis, when the computation is done on a computer.

From the foregoing it ranspires that the simplex algorithm involves the following sequence of steps. Let the constraint equations and the objective function be written in their canonical form. Then:

(1) evaluate $c_s=\min(c_j)$. If $c_s\geqslant 0$, the basic solution may be written as: $x_1=b_1$, $x_2=b_2$, ..., $x_m=b_m$, $x_{m+1}=0$, $x_{m+2}=0$, ..., $x_n=0$ (optimally), and $z=z_0$;

(2) if $c_s<0$, introduce x_s into the basis;

(3) ascertain if all $a_{is}\leqslant 0$. If so, the function z is not bounded below, because x_s can be increased without bound;

(4) if some $a_{is}>0$, find $b_m/a_{rs}=\min(b_i/a_{is})$ from among $a_{is}>0$. Eliminate x_r from the basis;

(5) find the new value of z_0

$$z_0^*=z_0+c_s b_r/a_{rs}$$

(6) transform the system of equations by substituting Eqs. (2.14) through (2.17) or Eqs. (2.9) in all the previous equations.

During the next step of optimization, steps (1) to (6) are repeated.

The flow diagram of the simplex algorithm is shown in Fig. 2.1.

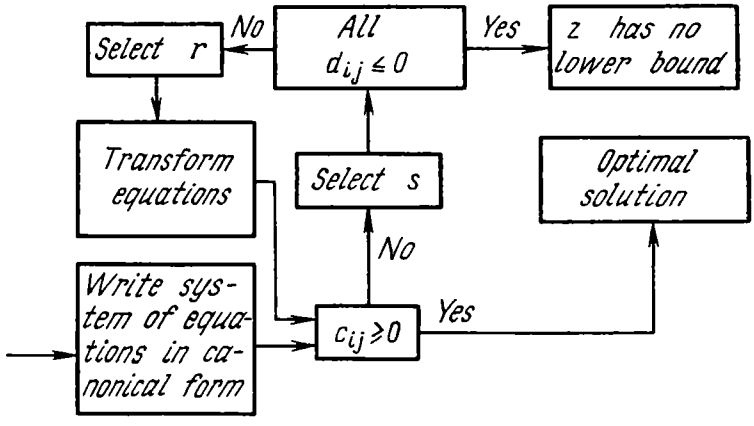


Fig. 2.1

Example 2.1. Let the original constraint equations and the objective function be written as

$$x_1-\frac{3}{4}x_2+2x_3-\frac{1}{4}x_4 \quad =3 \rightarrow *$$

$$x_5-\frac{1}{4}x_2+3x_3-\frac{3}{4}x_4 \quad =5$$

$$8x_2-24x_3+5x_4-z=-28$$

$$\uparrow$$

The first basic solution will be: $x_1=3$, $x_5=5$, $x_2=x_3=x_4=0$, $z=28$.

The first optimization step:

(1) select s: $c_s=\min(c_j)=-24$; $s=3$. Introduce the non-basic variable x_3 into the basis;

(2) evaluate the positive a_{is}'s: (2, 3);

(3) select r: $b_r/a_{rs}=\min(3/2, 5/3)=3/2$; $r=1$. Eliminate x_1 from the basis;

(4) evaluate x_3: $x_3=3/2-x_1/2+3x_2/8+x_4/8$ from the first row. Substitute

* A vertical arrow indicates introduction into, and a horizontal arrow, elimination from the basis.

the value of x_3 in the second and third equations to get a new system of equations:

$$x_3 + \frac{1}{2}x_1 - \frac{3}{8}x_2 - \frac{1}{8}x_4 = \frac{3}{2}$$

$$x_5 - \frac{3}{2}x_1 + \frac{7}{8}x_2 - \frac{3}{8}x_4 = \frac{1}{2} \rightarrow$$

$$12x_1 - x_2 + 2x_4 - z = 8$$

$$\uparrow$$

The new basic solution is: $x_3 = 3/2$, $x_5 = 1/2$, $x_1 = x_2 = x_4 = 0$, $z = -8$. The change in z is $c_s b_r / a_{rs} = -24 \times 3/2 = -36$.

The second optimization step:

(1) select s: $c_s = \min(c_j) = -1$; $s = 2$. Introduce x_2 into the basis;
(2) evaluate the positive coefficient in the s-column;
(3) select r: $r = 5$; eliminate x_5 from the basis;
(4) evaluate x_2 from the second line of equations:

$$x_2 = 4/7 - 8x_5/7 + 12x_1/7 + 3x_4/7$$

Substitute the value of x_2 in the first and third equation to get a new system of equations

$$x_2 - \frac{12}{7}x_1 - \frac{3}{7}x_4 + \frac{8}{7}x_5 = \frac{4}{7}$$

$$x_3 - \frac{1}{7}x_1 + \frac{2}{7}x_1 + \frac{3}{7}x_5 = \frac{12}{7}$$

$$\frac{72}{7}x_1 + \frac{11}{7}x_4 + \frac{8}{7}x_5 - z = \frac{60}{7}$$

The new basic solution is: $x_2 = 4/7$, $x_3 = 12/7$, $x_1 = x_4 = x_5 = 0$. The change in z is $c_s b_r / a_{rs} = -1\left(\frac{1}{2}\Big/\frac{7}{8}\right) = -4/7$.

Since all c_j in the objective function are positive, the basic solution thus obtained is optimal.

The following special cases of practical interest may be encountered in the use of the simplex algorithm.

1. If, apart from constraint equalities, there are constraint inequalities, the latter should be converted to equalities by adding an additional non-negative unknown. Let, for example, there be a constraint inequality of the form

$$2x_1 - 4x_2 + x_3 \leqslant 7$$

By adding a non-negative unknown, say x_6, we obtain an equality of the same meaning:

$$2x_1 - 4x_2 + x_3 + x_6 = 7; \qquad x_6 \geqslant 0$$

Thus, the inequality, $2x_1-4x_2+x_3\leqslant 7$, has been converted to an equality, $2x_1-4x_2+x_3+x_6=7$.

2. If any unknown may be negative, it should be replaced by the difference of two non-negative quantities, so as to avoid the violation of the constraint, Eq. (2.2):

$$x_2=x_7-x_8$$

where x_7 and x_8 are non-negative, and x_2 may take any sign.

3. If there are two unknowns for which the ratios b_i/a_{is} are minimal and identical, any r may be chosen for $b_i>0$.

4. If any one of the b_i' s is equal to zero, r is to be chosen from that line. However, $x_s^*=0$ and, as a consequence, z is not reduced in this optimization step.

5. If two of the b_i's are equal to zero, r may be chosen at will. However, the optimization will proceed at a faster pace if r be chosen for the greater value of a_{is}.

Example 2.2. Let there be an electric power system incorporating three stations with installed capacities P_1, P_2 and P_3 and with the fixed specific costs γ_1, γ_2 and γ_3. For simplicity, we assume them to be independent of the operating state of the system. The total output of the three stations has been specified in advance and is equal to W.

Find the hours of operation for each station, h_1, h_2 and h_3, that would minimize the cost, z:

$$z=P_1h_1\gamma_1+P_2h_2\gamma_2+P_3h_3\gamma_3$$

The unknowns h_1, h_2 and h_3 are subject to the following constraints:

$$h_1\leqslant 7000, \qquad h_2\leqslant 7000, \qquad h_3\leqslant 7000$$

By denoting $x_1=h_1$, $x_2=h_2$ and $x_3=h_3$, the constraints may be stated as

$$x_1\leqslant 7000, \qquad x_2\leqslant 7000, \qquad x_3\leqslant 7000$$

To convert the constraint inequalities to equalities, we add variables x_4, x_5 and x_6 to each, respectively. Now the constraint equations and the objective function take the form

$$\left.\begin{aligned} P_1x_1 + P_2x_2 + P_3x_3 &= W \\ x_1 + x_4 &= 7000 \\ x_2 + x_5 &= 7000 \\ x_3 + x_6 &= 7000 \\ P_1x_1\gamma_1+P_2x_2\gamma_2+P_3x_3\gamma_3 - z &= 0 \end{aligned}\right\} \qquad (2.18)$$

For conversion to the canonical form, we choose x_1, x_2, x_3 and x_6 as basic variables. Then x_4 and x_5 will be non-basic variables. To eliminate

the basic variables from the remaining equations, we can use the equalities derivable from Eq. (2.18):

$$x_1 = 7000 - x_4$$
$$x_2 = 7000 - x_5$$
$$x_3 = \frac{W - P_1 x_1 - P_2 x_2}{P_3} = \frac{W - P_1(7000 - x_4) - P_2(7000 - x_5)}{P_3}$$

Eliminate x_3 from all equations except the first; x_1 from all equations except the second, and x_2 from all equations except the third. This gives the following system of equations in canonical form

$$\begin{aligned}
x_3 - \frac{P_1}{P_3} x_4 - \frac{P_2}{P_3} x_5 &= \frac{W - P_1 7000 - P_2 7000}{P_3} \\
x_2 - x_5 &= 7000 \\
x_1 - x_4 &= 7000 \\
\frac{P_1}{P_3} x_4 + \frac{P_2}{P_3} x_5 + x_6 &= 7000 \frac{W - P_1 7000 - P_2 7000}{P_3}
\end{aligned}$$

$$P_1(\gamma_3 - \gamma_1)x_4 + P_2(\gamma_3 - \gamma_2)x_5 - z = -P_1\gamma_1 7000 - P_2\gamma_2 7000 - \gamma_3(W - P_1 7000 - P_2 7000)$$

The basic solution is: $x_1 = 7000$, $x_2 = 7000$ and $x_3 = (W - P_1 7000 - P_2 7000)/P_3$.

The coefficients c_j in the objective function are $P_1(\gamma_3 - \gamma_1)$ and $P_2(\gamma_3 - \gamma_2)$.

It follows then that if $\gamma_3 > \gamma_1$ and $\gamma_2 > \gamma_1$, the coefficients are positive and the solution is optimal. This is so because if the specific cost of station *3* exceeds that of stations *1* and *2*, the latter should be loaded to capacity (7000 h), and station *3* should be assigned utilization hours necessary to give the specified total output equal to $(W - P_1 7000 - P_2 7000)/P_3$.

If $\gamma_3 < \gamma_1$ and $\gamma_3 > \gamma_2$, it is necessary to introduce x_4 into the basis, that is, to reduce load on station *1* (as being less economical than station *3*) and eliminate x_6 from the basis, that is, to load station *3* to capacity.

2.4. Simplex Method

The simplex method is used to solve linear optimization problems in the general case, when the constraint equations and the objective function are in a form other than canonical. Then the system of equations has the form:

$$\left.\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \\
c_1x_1 + c_2x_2 + \ldots + c_nx_n - z &= 0
\end{aligned}\right\} \qquad (2.19)$$

The simplex method (SM) may be broken up into two stages: stage 1 is concerned with converting the constraint equations and the objective function to canonical form, which is not always possible (see below). Stage 2 seeks to optimize the objective function derived during stage 1, through the use of the simplex algorithm, SA (see Sec. 2.3).

Consider stage 1 of the SM.

At this stage, the system of equations is re-cast into canonical form by adding artificial variables to all equations, including all constraint equations. The only exception are the equations containing variables which are basic (that is, not appearing in any other equations and in the objective function) and also those with coefficients of unity. One more equation, the sum of all artificial variables, is added to the system. As a result, the system of equations takes the following form (in the general case):

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} &&= b_1\\
&a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x_{n+2} &&= b_2\\
&\ldots\ldots\ldots

and

$$W_0 = \sum_{i=1}^{m} b_i$$

the system of equations necessary to proceed with stage 1 of the simplex method finally takes the form

$$\left.\begin{array}{llll}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} & & & = b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & + x_{n+2} & & = b_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n & & + x_{n+m} & = b_m \\
c_1x_1 + c_2x_2 + \ldots + c_nx_n & & - z & = 0 \\
d_1x_1 + d_2x_2 + \ldots + d_nx_n & & & - W = - W_0
\end{array}\right\} \quad (2.20)$$

Instead of the z-function, stage 1 of the simplex method minimizes the W-function, using the simplex algorithm in its conventional form. The steps involved in this minimization are as follows:

(1) evaluate the d_j's. If all $d_j \geqslant 0$, W cannot be minimized. If $W > 0$, there is no feasible solution, except for $W = W_0$.

If some $d_j < 0$, the index s of the variable introduced into the basis is chosen from the condition

$$d_s = \min (d_j)$$

where $d_s < 0$;

(2) determine the index r of the unknown eliminated from the basis from the condition

$$b_r/a_{rs} = \min (b_i/a_{is})$$

where $a_{is} > 0$;

(3) convert all equations of the system, Eqs. (2.20), using equations (2.14) through (2.17), and also the following equations for d_j and W_0:

for all columns except r

$$d_j^* = d_j - d_s a_{rj}/a_{rs} \quad (2.21)$$

for the r-column

$$\begin{aligned} d_r^* &= - d_s/a_{rs} \\ W_0^* &= W_0 + d_s b_r/a_{rs} \end{aligned} \quad (2.22)$$

Then steps (1) through (3) are repeated until all quantities become non-negative;

(4) evaluate W. If $W = 0$, all artificial variables will obviously be equal to zero. Then re-write Eqs. (2.20) so as to eliminate the last equation and all artificial variables. The resultant system will be in canonical form. If $W > 0$, there is no solution.

Stage 2 of the simplex method optimizes the system yielded by stage 1, by means of the simplex algorithm.

The flow diagram of the simplex method is shown in Fig. 2.2.

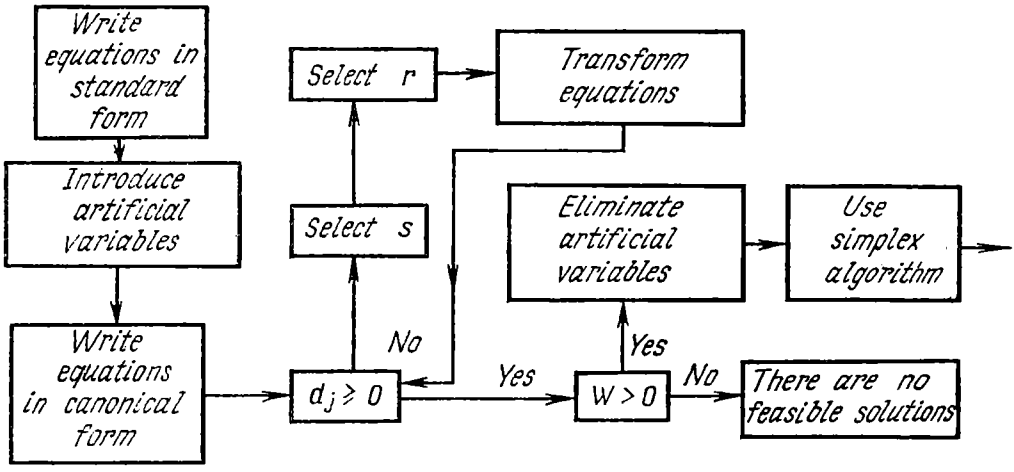


Fig. 2.2

Example 2.3. Let there be a system of equations

$$\begin{aligned} 5x_1-4x_2+13x_3-2x_4+x_5 &= 20\\ x_1-x_2+5x_3+x_4+x_5 &= 8\\ x_1+6x_2-7x_3+x_4+5x_5-z &= 0 \end{aligned}$$

Add artificial variables x_6 and x_7 and an equation for W:

$$x_6+x_7-W=0$$

Eliminate x_6 and x_7 by subtracting the sum of the first two equations of the system. This gives a system of equations in canonical form:

$$\begin{aligned} 5x_1-4x_2+13x_3-2x_4+x_5+x_6 &= 20\\ x_1-x_2+5x_3-x_4+x_5+x_7 &= 8\\ x_1+6x_2-7x_3+x_4+5x_5-z &= 0\\ -6x_1+5x_2-18x_3+3x_4-2x_5-W &= -28 \end{aligned}$$

The first optimization step:

select s: $d_s=\min d_j=-18$; $s=3$. Introduce x_3 into the basis;

select r: $b_r/a_{rs}=\min(b_i/a_{is})=\min(20/13, 8/5)=20/13$; $r=6$. Eliminate x_6 from the basis.

After conversion by Eqs. (2.14) through (2.17) and (2.21), we obtain the following system of equations

$$\begin{aligned} x_3+\frac{5}{13}x_1-\frac{4}{13}x_2-\frac{2}{13}x_4+\frac{1}{13}x_5+\frac{1}{13}x_6 &= \frac{20}{13}\\ -\frac{12}{13}x_1+\frac{7}{13}x_2-\frac{3}{13}x_4+\frac{8}{13}x_5-\frac{5}{13}x_6+x_7 &= \frac{4}{13}\\ \frac{48}{13}x_1+\frac{50}{13}x_2-\frac{1}{13}x_4+\frac{72}{13}x_5+\frac{7}{13}x_6-z &= \frac{140}{13}\\ \frac{12}{13}x_1-\frac{7}{13}x_2+\frac{3}{13}x_4-\frac{8}{13}x_5+\frac{8}{13}x_6-W &= -\frac{4}{13} \end{aligned}$$

The second optimization step:

select s: $d_s=\min d_j=-8/13$; $s=5$. Introduce x_5 into the basis;

select r: $b_r/a_{rs}=\min(20, 1/2)=1/2$; $r=7$. Eliminate x_7 from the basis. After conversion, we get the following system:

$$\begin{array}{rcl} x_3+\frac{1}{2}x_1-\frac{3}{8}x_2-\frac{1}{8}x_4+\frac{1}{8}x_6-\frac{1}{8}x_7 & & =\frac{3}{2} \\ x_5\quad -\frac{3}{2}x_1+\frac{7}{8}x_2-\frac{3}{8}x_4-\frac{5}{8}x_6+\frac{13}{8}x_7 & & =\frac{1}{2} \\ 12x_1-\ x_2+2x_4+4x_6-9x_7 & -z & =8 \\ x_6+\ x_7 & -W & =0 \end{array}$$

Since $W=0$, it cannot be reduced any longer. Thus, $x_6=x_7=0$.

Stage 1 terminates by eliminating the columns x_6 and x_7 and the last equation and producing the following system of equations:

$$\begin{array}{rcl} x_3+\frac{1}{2}x_1-\frac{3}{8}x_2-\frac{1}{8}x_4 & & =\frac{3}{2} \\ x_5\quad -\frac{3}{2}x_1+\frac{7}{8}x_2+\frac{3}{8}x_4 & & =\frac{1}{2} \\ 12x_1-\ x_2+2x_4 & -z & =8 \end{array}$$

Stage 2 uses the simplex algorithm to minimize z.

Example 2.4. It is planned to build three electric power stations, *1*, *2* and *3*. Their maximum capacities are set at $P_{1\,\text{lim}}=900$ MW, $P_{2\,\text{lim}}=1200$ MW and $P_{3\,\text{lim}}=1200$ MW. The operating hours of the stations are constrained to be $3000\leqslant h_i\leqslant 7000$. There are also constraints on the balance of power and energy:

$$\sum_{i=1}^{3} P_i=2600 \text{ MW}$$

and

$$\sum_{i=1}^{3} P_i h_i=15.6 \text{ TWh}$$

The objective function is

$$z=\sum_{i=1}^{3}(a_i+b_i h_i)P_i$$

where a_i = discounted investment costs including renovation and repair, roubles/kW

b_i = discounted fuel costs per MWh for each station

h_i = operating hours of each station

We assume that the investment costs and the costs b_i are linear functions of h_i. Then

$$b_i=b_i^{\min}+\frac{b_i^{\max}-b_i^{\min}}{h_i^{\max}-h_i^{\min}}(h_i^{\max}-h_i)$$

It can be proved that any station operating h_i hours at per MWh cost b_i has to meet higher costs than two identical stations operating $h_i^{\min}$ and $h_i^{\max}$

hours with capacities $P_i^{(1)}$ and $P_i^{(2)}$, respectively, where

$$P_i = P_i^{(1)} + P_i^{(2)}$$

$$h_i = \frac{h_i^{\min} P_i^{(1)} + h_i^{\max} P_i^{(2)}}{P_i^{(1)} + P_i^{(2)}} = \frac{h_i^{\min} P_i^{(1)} + h_i^{\max} P_i^{(2)}}{P_i}$$

The reduction in cost is

$$m = [P_i^{(1)} + P_i^{(2)}] h_i b_i - [P_i^{(1)} h_i^{\min} b_i^{\max} + P_i^{(2)} h_i^{\max} b_i^{\min}]$$

Substituting h_i and b_i in the above equation and carrying out some manipulations, we get:

$$m = \frac{P_i^{(1)} P_i^{(2)}}{P_i^{(1)} + P_i^{(2)}} (h_i^{\max} - h_i^{\min})(b_i^{\max} - b_i^{\min}) > 0$$

that is, the break-down of a station into two identical ones will cut down the costs. In this case, the objective function will have the form

$$z = \sum_{i=1}^{3} [(a_i + b_i^{\max} h_i^{\min}) P_i^{(1)} + (a_i + b_i^{\min} h_i^{\max}) P_i^2]$$

This function is equivalent to quantifying a single duration curve into two levels representing operation for 3000 and 7000 hours, respectively, to give the same annual generation. Suppose that all terms in parentheses are evaluated and the objective function is

$$z = 15P_1^{(1)} + 20P_1^{(2)} + 20P_2^{(1)} + 25P_2^{(2)} + 20P_3^{(1)} + 30P_3^{(2)}$$

subject to

$$P_1^{(1)} + P_1^{(2)} \leqslant 900; \qquad P_2^{(1)} + P_2^{(2)} \leqslant 1200; \qquad P_3^{(1)} + P_3^{(2)} \leqslant 1200$$

$$P_1^{(1)} + P_1^{(2)} + P_2^{(1)} + P_2^{(2)} + P_3^{(1)} + P_3^{(2)} = 2600$$

$$3P_1^{(1)} + 7P_1^{(2)} + 3P_2^{(1)} + 7P_2^{(2)} + 3P_3^{(1)} + 7P_3^{(2)} = 15{,}600$$

We denote $P_1^{(1)} = x_1$, $P_1^{(2)} = x_2$, $P_2^{(1)} = x_3$, $P_2^{(2)} = x_4$, $P_3^{(1)} = x_5$ and $P_3^{(2)} = x_6$. To convert the inequalities to equalities we add the unknowns x_7, x_8, and x_9, and to re-cast the equations to their canonical form, we add x_{10} and x_{11}. The computations are summarized in the accompanying tables where the basic variables are represented by zeros.

Stage 1 of the simplex method

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	$-z$	$-W$	b
1	1	1	1	1	1	0	0	0	1	0	0	0	2600
3	7	3	7	3	7	0	0	0	0	1	0	0	15600
1	1	0	0	0	0	1	0	0	0	0	0	0	900→
0	0	1	1	0	0	0	1	0	0	0	0	0	1200
0	0	0	0	1	1	0	0	1	0	0	0	0	1200
15	20	20	25	20	30	0	0	0	0	0	1	0	0
−4	−8	−4	−8	−4	−8	0	0	0	0	0	0	1	−18200

Introduce x_2 into, and remove x_7 from, the basis. After conversion, we get:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	$-z$	$-W$	b
0	0	1	1	1	1	−1	0	0	1	0	0	0	1700
−4	0	3	7	3	7	−7	0	0	0	1	0	0	9300
1	1	0	0	0	0	1	0	0	0	0	0	0	900
0	0	1	1	0	0	0	1	0	0	0	0	0	1200→
0	0	0	0	1	1	0	0	1	0	0	0	0	1200
−5	0	20	25	20	30	−20	0	0	0	0	1	0	−18000
4	0	−4	−8	−4	−8	8	0	0	0	0	0	1	−11000
			↑										

Introduce x_4 into, and remove x_8 from, the basis:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	$-z$	$-W$	b
0	0	0	0	1	1	−1	−1	0	1	0	0	0	500
−4	0	−4	0	3	7	−7	−7	0	0	1	0	0	900
1	1	0	0	0	0	1	0	0	0	0	0	0	900
0	0	1	1	0	0	0	1	0	0	0	0	0	1200
0	0	0	0	1	1	0	0	1	0	0	0	9	1200
−5	0	−5	0	20	30	−20	−25	0	0	0	1	0	−48000
4	0	4	0	−4	−8	8	8	0	0	0	0	1	−1400

Introduce x_6 into, and remove x_{11} from, the basis:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	$-z$	$-W$	b
$\frac{4}{7}$	0	$\frac{4}{7}$	0	$\frac{4}{7}$	0	0	0	0	1	$-\frac{1}{7}$	0	0	$\frac{2600}{7}$→
$-\frac{4}{7}$	0	$-\frac{4}{7}$	0	$\frac{3}{7}$	1	−1	−1	0	0	$\frac{1}{7}$	0	0	$\frac{900}{7}$
1	1	0	0	0	0	1	0	0	0	0	0	0	900
0	0	1	1	0	0	0	1	0	0	0	0	0	1200
$\frac{4}{7}$	0	$\frac{4}{7}$	0	$\frac{4}{7}$	0	1	1	1	0	$-\frac{1}{7}$	0	0	$\frac{7500}{7}$
$\frac{85}{7}$	0	$\frac{85}{7}$	0	$\frac{50}{7}$	0	10	5	0	0	$\frac{30}{7}$	1	0	$-\frac{363000}{7}$
$-\frac{4}{7}$ ↑	0	$-\frac{4}{7}$	0	$-\frac{4}{7}$	0	0	0	0	0	$\frac{8}{7}$	0	1	$-\frac{2600}{7}$

On introducing x_1 into, and eliminating x_{10} from, the basis:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	$-z$	$-W$	b
1	0	1	0	1	0	0	0	0	$\frac{7}{4}$	$-\frac{1}{4}$	0	0	650
0	0	0	0	1	1	−1	−1	0	1	0	0	0	500
0	1	−1	0	−1	0	1	0	0	$-\frac{7}{4}$	$\frac{1}{4}$	0	0	250
0	0	1	1	0	0	0	1	0	0	0	0	0	1200
0	0	0	0	0	0	1	1	1	1	−1	0	0	700
0	0	0	0	−5	0	10	5	0	$-\frac{85}{4}$	$-\frac{5}{4}$	1	0	−59750
0	0	0	0	0	0	0	0	0	1	1	0	1	0

Stage 2 of the simplex method

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	$-z$	b
1	0	1	0	1	0	0	0	0	0	650
0	0	0	0	1	1	−1	−1	0	0	500→
0	1	−1	0	−1	0	1	0	0	0	250
0	0	1	1	0	0	0	1	0	0	1200
0	0	0	0	0	0	1	1	1	0	700
0	0	0	0	−5	0	10	5	0	1	−59750
				↑						

On introducing x_5 into, and eliminating x_6 from, the basis:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	$-z$	b
1	0	1	0	0	−1	1	1	0	0	150
0	0	0	0	1	1	−1	−1	0	0	500
0	−1	−1	0	0	1	0	−1	0	0	750
0	0	1	1	0	0	0	1	0	0	1200
0	0	0	0	0	0	1	1	1	0	700
0	0	0	0	0	5	5	0	1	1	−58250

Since $W=0$, and also $x_{10}=0$ and $x_{11}=0$, stage 1 is terminated. We delete the columns for x_{10}, x_{11} and for $-W$, and also the last row.

Because all the c_j's are greater than zero, the solution obtained is an optimum one:

$$P_1=900, \quad h_1=(150\times 3000+750\times 7000)/900=6333$$

$$P_2=1200, \quad h_2=7000; \qquad P_3=500, \quad h_3=3000; \qquad z=58{,}250$$

The solution is approximate not only because we have assumed that the investment costs are linear, but also because we have approximated the duration curve with two levels, and have assumed the dependence $b_{i\,opt}\,h$ to be linear.

The result would be more accurate, if the duration curve were approximated by s levels. Then,

$$z=\sum_{i=1}^{n}\sum_{l=1}^{s}(a_i+b_{il}h_l)P_{il}$$

where l is the level index and n is the number of stations.

The values of a_i, b_{il} and h_l are specified in advance. The additional constraints are:

$$\sum_{l=1}^{s}P_{il}=P_i$$

$$\sum_{i=1}^{n}=P_{il}=P_l$$

where P_l is the capacity for level l. The problem is solved by the simplex method.

2.5. Modified Simplex Method

A limitation of the conventional simplex method examined in Sec. 2.4 is that when we convert the constraint equations and those for z and W in passing to the next optimization step, we have to re-calculate all of the coefficients and free terms of the complete system of equations, although we need only know the coefficients of the W-equation or of the z-equation (according to the stage of the simplex method) in order to select s, and only the coefficients of the x_s column and free terms in order to select r. We are forced to do so because in the simplex method the sought coefficients of the $(k+1)$st step are computed from those of the kth step. This adds to the burden of computational work and slows down the search for an optimum solution.

The rationale of the modified simplex method (MSM) is that instead of all, we evaluate only the coefficients actually needed, and do so directly from the original system of equations (the zeroth cycle) rather than from the data of the previous cycle.

This form of computation is possible, if we introduce new matrix quantities which can be defined for any, say kth, cycle as follows:

(1) a square matrix which is the inverse basis of the kth cycle:

$$\mathbf{B}_k^{-1}=\begin{vmatrix}\beta_{11} & \beta_{12} & \cdots & \beta_{1m}\\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mm}\end{vmatrix} \tag{2.23}$$

(2) the column vectors π_k and σ_k of the simplex coefficients of the kth cycle:

$$\pi_k = \begin{vmatrix} \pi_{1k} \\ \pi_{2k} \\ \cdots \\ \pi_{mk} \end{vmatrix}, \qquad \sigma_k = \begin{vmatrix} \sigma_{1k} \\ \sigma_{2k} \\ \cdots \\ \sigma_{mk} \end{vmatrix} \tag{2.24}$$

The basis $\mathbf{B}_k$ of the kth cycle is the coefficient matrix of the constraint equations in the zeroth cycle for the unknowns which are to become basic variables in the kth cycle, that is, for which the coefficient matrix in the kth cycle will be a unit matrix. $\mathbf{B}_k$ cannot be evaluated in advance, because we do not know which variables will be the basic ones in the kth cycle.

The *inverse basis*, $\mathbf{B}_k^{-1}$, of the kth cycle is a matrix which is the inverse of $\mathbf{B}_k$. The zeroth-cycle basic variables in the zeroth-cycle system of equations have a unit matrix $\mathbf{E}$. It can be shown that the matrix of the zeroth-cycle basic variables will be equal to $\mathbf{B}_k^{-1}$ in the kth cycle.

Consider the constraint equation for the zeroth cycle in contracted matrix notation

$$\mathbf{A}_{ij}x_j = b_i \tag{2.25}$$

where $\mathbf{A}_{ij}$ is the coefficient matrix of the constraint equation and b_i is the column vector of the free terms.

Let $\mathbf{A}'_{ij}$ be the coefficient submatrix corresponding to the zeroth-cycle basic unknowns, that is, $\mathbf{A}'_{ij} = \mathbf{E}$. Also, let $\mathbf{A}''_{ij}$ be the coefficient submatrix corresponding to the kth-cycle basic unknowns in the zeroth cycle. By definition, $\mathbf{A}''_{ij} = \mathbf{B}_k$.

Multiplication (say, post-multiplication) of the zeroth-cycle constraint equation in matrix form by $\mathbf{B}_k^{-1}$ reduces the coefficient submatrix of the zeroth-cycle basic variables to $\mathbf{E}\mathbf{B}_k^{-1} = \mathbf{B}_k^{-1}$, and the coefficient submatrix of the kth-cycle basic variable to $\mathbf{B}_k\mathbf{B}_k^{-1} = \mathbf{E}$. That is, it becomes a unit matrix corresponding to the basic variables of the kth cycle.

Thus, in order to evaluate $\mathbf{B}_k^{-1}$ of the kth cycle, we need only know the coefficients in the columns for the basic variables of the zeroth cycle. This is why in the modified simplex method it is important to note changes in the coefficients *solely* in those columns. When all coefficients in these columns have been evaluated in the kth cycle, the inverse basis $\mathbf{B}_k^{-1}$ of the kth cycle is evaluated as the coefficient submatrix for the basic variables of the zeroth cycle. $\mathbf{B}_k^{-1}$ is re-evaluated consecutively from cycle to cycle, starting with the zeroth cycle (see below).

The simplex coefficients π_k of the kth cycle have the property that if we *multiply* the first equation in the zeroth-cycle system of constraint equations by π_{1k}, the second equation by π_{2k}, etc., and *subtract* the sum of these equations from the z-equation in the zeroth cycle, the result will be the z-equation for the kth cycle, from which all basic unknowns of the kth cycle have been eliminated, that is, they will occur with zero coefficients.

Since in the zeroth-cycle the basic unknowns have a unit coefficient matrix, in the kth cycle their coefficients in the z-equation are $\pi_{1k}, \pi_{2k}, \ldots$.

To state this differently, if we manage to evaluate in some manner the coefficients c_j in the kth cycle for the basic variables of the zeroth cycle, these coefficients taken with the opposite sign will be the sought simplex coefficients π_{jk} of the kth cycle.

The simplex coefficients σ_k are analogous to the coefficients π_k, but they occur in the W-equation (stage 1 of the simplex method), and not in the z-equation.

If we evaluate the coefficients d_j in the kth cycle for the basic variables of the zeroth cycle, these coefficients taken with the opposite sign will be the sought simplex coefficients σ_{jk} of the kth cycle.

The coefficients π_{jk} and σ_{jk} are re-calculated in the same manner as $\mathbf{B}_k^{-1}$, from cycle to cycle, starting with the zeroth one.

Thus, as we move from cycle to cycle, we need only re-calculate $\mathbf{B}_k^{-1}$, π_k and σ_k. The remaining coefficients and free terms required for computation are evaluated for the kth cycle directly from the data of the zeroth cycle, which is a distinction and advantage of the modified simplex method.

Design equations of the MSM. If $\mathbf{B}_k^{-1}$, π_k and σ_k are known, all coefficients a_{ij} and free terms b_i for the kth cycle can be evaluated as follows.

In matrix notation, the constraint equation of the zeroth cycle may be written as

$$\mathbf{A}_{ij}^{(0)} x_j = b_i^{(0)}$$

where

$$\mathbf{A}_{ij}^{(0)} = \begin{vmatrix} a_{11}^{(0)} & a_{12}^{(0)} \ldots a_{1n}^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} \ldots a_{2n}^{(0)} \\ \ldots\ldots\ldots\ldots \\ a_{m1}^{(0)} & a_{m2}^{(0)} \ldots a_{mn}^{(0)} \end{vmatrix}, \quad x_j = \begin{vmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{vmatrix}, \quad b_i^{(0)} = \begin{vmatrix} b_1^{(0)} \\ b_2^{(0)} \\ \ldots \\ b_n^{(0)} \end{vmatrix}$$

where the i's are the indexes of the rows corresponding to the basic variables of the kth cycle and the j's are the column indexes.

Premultiplying the matrix equation by $\mathbf{B}_k^{-1}$, that is, by

$$\begin{vmatrix} \beta_{11}^{(k)} & \beta_{12}^{(k)} \ldots \\ \beta_{21}^{(k)} & \beta_{22}^{(k)} \ldots \\ \ldots\ldots\ldots\ldots \end{vmatrix}$$

gives $\mathbf{B}_k^{-1} A_{ij}^{(0)} x_j = B_k^{-1} b_i^{(0)}$. However, $A_{ij}^{(k)} x_j = b_i^{(k)}$.

Hence

$$A_{ij}^{(k)} = B_k^{-1} A_{ij}^{(0)} = \begin{vmatrix} \beta_{11}^{(k)} & \beta_{12}^{(k)} \ldots \\ \ldots\ldots\ldots\ldots \\ \beta_{i1}^{(k)} & \beta_{i2}^{(k)} \ldots \\ \ldots\ldots\ldots\ldots \end{vmatrix} \cdot \begin{vmatrix} a_{11}^{(0)} & a_{12}^{(0)} \ldots a_{1j}^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} \ldots a_{2j}^{(0)} \end{vmatrix}$$

whence

$$a_{ij}^{(k)}=\beta_{i1}^{(k)}a_{1j}^{(0)}+\beta_{i2}^{(k)}a_{2j}^{(0)}+\ldots=\sum_{l=1}^{m}\beta_{il}^{(k)}a_{lj}^{(0)} \tag{2.26}$$

Since

$$b^{(k)}=\mathbf{B}_k^{-1}b^{(0)}=\begin{vmatrix}\beta_{11}^{(k)} & \beta_{12}^{(k)} \ldots\\ \beta_{21}^{(k)} & \beta_{22}^{(k)} \ldots\\ \ldots\ldots\ldots\\ \beta_{i1}^{(k)} & \beta_{i2}^{(k)}\end{vmatrix}\cdot\begin{vmatrix}b_1^{(0)}\\ b_2^{(0)}\\ \ldots\\ b_m^{(0)}\end{vmatrix} \tag{2.27}$$

then

$$b_i^{(k)}=\beta_{i1}^{(k)}b_1^{(0)}+\beta_{i2}^{(k)}b_2^{(0)}+\ldots=\sum_{l=1}^{m}\beta_{il}^{(k)}b_l^{(0)}$$

Next, by the definition of π_{jk} and σ_{jk}, we get

$$c_j^{(k)}=c_j^{(0)}-\pi_{1k}a_{1j}^{(0)}-\pi_{2k}a_{2j}^{(0)}-\ldots=c_j^{(0)}-\sum_{l=1}^{m}\pi_{lk}a_{lj}^{(0)} \tag{2.28}$$

$$d_j^{(k)}=d_j^{(0)}-\sigma_{1k}a_{1j}^{(0)}-\sigma_{2k}a_{2j}^{(0)}-\ldots=d_j^{(0)}-\sum_{l=1}^{m}\sigma_{lk}a_{lj}^{(0)} \tag{2.29}$$

$$z_0^{(k)}=z_0^{(0)}+\pi_{1k}b_1^{(0)}+\pi_{2k}b_2^{(0)}+\ldots=z_0^{(0)}+\sum_{l=1}^{m}\pi_{lk}b_l^{(0)} \tag{2.30}$$

$$W_0^{(k)}=W_0^{(0)}+\sigma_{1k}b_1^{(0)}+\sigma_{2k}b_2^{(0)}+\ldots=W_0^{(0)}+\sum_{l=1}^{m}\sigma_{lk}b_l^{(0)} \tag{2.31}$$

Equations (2.26) through (2.31) can be used to evaluate all coefficients and free terms of the system of equations for the kth cycle, except the coefficients of the zeroth-cycle basic unknowns.

$\mathbf{B}_k^{-1}$, π_{jk} and σ_{jk} are re-calculated from cycle to cycle upon the selection of s and r as follows.

For all rows except the s-row [see Eqs. (2.14) through (2.16)]:

$$\beta_{ij}^{*}=\beta_{ij}-a_{is}\beta_{rj}/a_{rs} \tag{2.32}$$

For the s-row

$$\beta_{sj}^{*}=\beta_{rj}/a_{rs} \tag{2.33}$$

If it is introduced into the basis for the zeroth cycle, for the r-column

$$\beta_{ir}^{*}=a_{is}/a_{rs} \tag{2.34}$$

$$\beta_{sr}^{*}=1/a_{rs} \tag{2.35}$$

According to Eq. (2.17), we get

$$\pi_j^{*}=\pi_j+c_s\beta_{rj}/a_{rs} \tag{2.36}$$

$$\sigma_j^{*}=\sigma_j+d_s\beta_{rj}/a_{rs} \tag{2.37}$$

Here, the asterisk (*) is associated with the kth cycle; an unsubscripted quantity is associated with the $(k-1)$st cycle, the i's are the indexes of the

basic variables in the kth cycle, and the j's are the indexes of the basic variables in the zeroth cycle.

Equations (2.32) through (2.37) may be used to evaluate $\mathbf{B}_k^{-1}$, π_{jk} and σ_{jk} from the data of the previous, that is $(k-1)$st, cycle.

Example 2.5. The original system of equations in standard form is

$$\begin{aligned} 5x_1-4x_2+13x_3-2x_4+x_5 &= 20\\ x_1-x_2+5x_3-x_4+x_5 &= 8\\ x_1+6x_2-7x_3+x_4-z &= 0 \end{aligned}$$

After adding artificial variables, the coefficients and free terms associated with the zeroth cycle may be tabulated as follows:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	$-z$	$-W$	b
5	−4	13	−2	1	1				20
1	−1	5	−1	1		1			8
1	6	−7	1	5			1		0
−6	5	−18	3	−2				1	−28

In the zeroth cycle, the basic variables are x_6 and x_7. Thus, for the zeroth cycle,

$$\mathbf{B}_k^{-1}=\begin{vmatrix} 1 & \\ & 1 \end{vmatrix}, \quad \begin{aligned} &\beta_{66}=1; \quad \beta_{67}=0\\ &\beta_{76}=0, \quad \beta_{77}=1 \end{aligned}$$

$$\pi_{6k}=\pi_{7k}=\sigma_{6k}=\sigma_{7k}=0$$

Step one. The d_j row of coefficients may be written as:

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7
d_j	−6	5	−18	3	−2	0	0

Select s: $s=3$. Introduce x_3 into the basis.

Select r: $b_r/a_{rs}=\min(20/13, 8/5)=20/13$; $r=6$. Eliminate x_6 from the basis. Obviously, $a_{rs}=a_{63}=13$.

Evaluate $\mathbf{B}_k^{-1}$ [see Eqs. (2.32) through (2.35)].

The new designations of the x_3 and x_7 rows correspond to the new basic variables:

$$\beta_{36}^*=\beta_{66}/a_{63}=1/13$$

$$\beta_{37}^*=\beta_{67}/a_{63}=0/13=0$$

$$\beta_{76}^*=\beta_{76}-a_{73}\beta_{66}/a_{63}=0-5\times 1/13=-5/13$$

$$\beta_{77}^*=\beta_{77}-a_{73}\beta_{67}/a_{63}=1-5\times 0/13=1$$

Evaluate π_6, π_7, σ_6 and σ_7, by Eqs. (2.36) and (2.37):

$$\pi_6^*=\pi_6+c_3\beta_{66}/a_{63}=0+(-7)\times 1/13=-7/13$$
$$\pi_7^*=\pi_7+c_3\beta_{67}/a_{63}=0+(-7)\times 0/13=0$$
$$\sigma_6^*+\sigma_6+d_3\beta_{66}/a_{63}=0+(-18)\times 1/13=-18/13$$
$$\sigma_7^*=\sigma_7+d_3\beta_{67}/a_{63}=0+(-18)\times 0/13=0$$

The new values of $\mathbf{B}_k^{-1}$, π_k and σ_k can be tabulated as follows:

	x_6	x_7		x_6	x_7
x_3	1/13	0	$-\pi_k$	7/13	0
x_7	−5/13	1	$-\sigma_k$	18/13	0

Then,

$$W_0^{(k)}=28-18\times 20/13-0.8=-4/13$$

Step two. Find the new values of the d_j's by Eq. (2.29) and tabulate them:

$$d_1=-6+18\times 5/13=12/13$$
$$d_2=-6+18(-4)/13=-7/13$$
$$d_3=0$$
$$d_4=3+18(-2)/13=3/13$$
$$d_5=-2+18(1)/13=-8/13$$
$$d_6=0+18(1)/13=18/13$$
$$d_7=0$$

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7
d_j	12/13	−7/13	0	3/13	−8/13	18/13	0

Select s: $s=5$. Introduce x_5 into the basis. Before we can select r, it is necessary to evaluate $a_{35}^{(k)}$, $a_{75}^{(k)}$, $b_3^{(k)}$ and $b_7^{(k)}$ by Eqs. (2.26) and (2.27):

$$a_{35}^{(k)}=1\times 1/13+0\times 1=1/13$$
$$a_{75}^{(k)}=-5\times 1/13+1\times 1=8/13$$
$$b_3^{(k)}=1\times 20/13+0\times 8=20/13$$
$$b_7^{(k)}=-5\times 20/13+1\times 8=4/13$$

Select r: $\min(20, 4/8)=4/8$; $r=7$. Eliminate x_7 from the basis; $a_{r6}=8/13$. Evaluate $\mathbf{B}_k^{-1}$, π_k and σ_k.

The new rows correspond to the basic variables x_3 and x_5:

$$\beta_{36}^{*}=\frac{1}{13}-\frac{1}{13}\left(\frac{-5/13}{8/13}\right)=\frac{1}{8}; \qquad \beta_{56}^{*}=\frac{-5/13}{8/13}=-\frac{5}{8}$$

$$\beta_{37}^{*}=0-\frac{1}{13}\left(\frac{1}{8/13}\right)=-\frac{1}{8}; \qquad \beta_{57}^{*}=\frac{1}{8/13}=\frac{13}{8}$$

$$\pi_6^{*}=-\frac{7}{13}+\frac{72}{13}\left(\frac{-5/13}{8/13}\right)=-4; \qquad \pi_7^{*}=0+\frac{72}{13}\left(\frac{1}{8/13}\right)=9$$

$$\sigma_6^{*}=-\frac{18}{13}+\left(\frac{-8}{13}\right)\left(\frac{-5/13}{8/13}\right)=-1; \qquad \sigma_7^{*}=0+\left(-\frac{8}{13}\right)\left(\frac{1}{8/13}\right)=-1$$

The new values of $\mathbf{B}_k^{-1}$, π_k and σ_k are then tabulated as follows:

	x_6	x_7		x_6	x_7
x_3	1/8	−1/8	$-\pi_k$	4	−9
x_5	−5/8	13/8	$-\sigma_k$	1	1

Evaluate $W_0^{(k)}$ by Eq. (2.31)

$$W_0^{(k)}=28-20-8=0$$

This completes stage 1. Here $x_6=x_7=0$.
To start stage 2, we define the row for all the c_j's [see Eq. (2.28)].

Step one. Find the new values of the c_j's and tabulate them:

$$c_1=1+4\times 5-9\times 1=12$$
$$c_2=6+4(-4)-9(-1)=-1$$
$$c_3=-7+4(13)-9(5)=0$$
$$c_4=1+4(-2)-9(-1)=2$$
$$c_5=5+4(1)-9(1)=0$$

x	x_1	x_2	x_3	x_4	x_5
c_j	12	−1	0	2	0

Select s: $s=2$. Introduce x_2 into the basis. In order to select r, we evaluate $a_{32}^{(k)}$, $a_{52}^{(k)}$, $b_3^{(k)}$ and $b_5^{(k)}$ [see Eq. (2.26) and (2.27)]:

$$a_{32}^{(k)}=-3/8; \qquad a_{52}^{(k)}=7/8; \qquad b_3^{(k)}=3/2; \qquad b_5^{(k)}=1/2$$

Select r: since only $a_{52}^{(k)}$ is greater than zero, then $r=5$. Eliminate x_5 from the basis. Evaluate $\mathbf{B}_k^{-1}$ and π_k. The new rows correspond to the basic variables x_2 and x_3:

$$\beta_{26}^*=\frac{-5/8}{7/8}=-\frac{5}{7}; \qquad \beta_{27}^*=\frac{13/8}{7/8}=\frac{13}{7}$$

$$\beta_{36}^*=-\frac{1}{8}-\left(-\frac{3}{8}\right)\frac{-5/8}{7/8}=-\frac{1}{7}; \qquad \beta_{37}^*=-\frac{1}{8}-\left(-\frac{3}{8}\right)\frac{13/8}{7/8}=\frac{4}{7}$$

$$\pi_6=-4+(-1)\frac{-5/8}{7/8}=-\frac{23}{7}; \qquad \pi_7^*=9+(-1)\frac{13/8}{7/8}=\frac{50}{7}$$

Tabulate the new values:

	x_6	x_7
x_2	$-5/7$	$13/7$
x_3	$-1/7$	$14/7$
$-\pi_k$	$23/7$	$-50/7$

Step two: Find the row for the c_j's [see Eq. (2.28)]:

$$c_1=72/7, \quad c_2=0, \quad c_3=0, \quad c_4=11/7, \quad c_5=8/7$$

Since none of the c_j's is negative, the optimization may be taken as being complete.

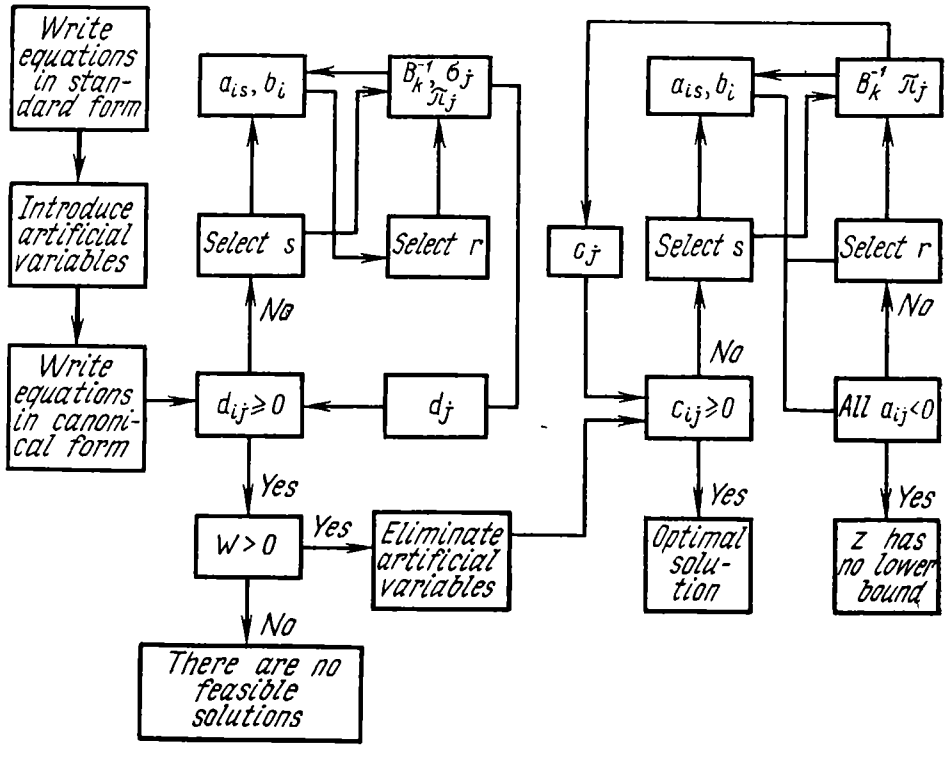


Fig. 2.3

The optimum solution is obtained by evaluating $b_2^{(k)}$, $b_3^{(k)}$ and $z_0^{(k)}$:

$$b_2^{(k)} = -5 \times 20/7 + 13 \times 8/7 = 4/7$$

$$b_3^{(k)} = -1 \times 20/7 + 4 \times 8/7 = 12/7$$

$$z_0^{(k)} = 0 - 23 \times 20/7 + 50 \times 8/7 = -60/7$$

The solution of the problem is: $x_2 = 4/7$, $x_3 = 12/7$, $x_1 = x_4 = x_5 = 0$, $z = -60/7$.

From an analysis of the solution obtained by the modified simplex method, we note that the following quantities were not computed during the second step of stage one:

$$a_{31}, \quad a_{32}, \quad a_{34}, \quad a_{71}, \quad a_{74}, \quad c_1, \quad c_2, \quad c_3, \quad c_4, \quad c_5$$

During the first step of stage two, the following quantities were not evaluated:

$$a_{31}, \quad a_{34}, \quad a_{35}, \quad a_{51}, \quad a_{53}, \quad a_{50}$$

Since they have to be evaluated in the conventional simplex method, this is a proof that the modified simplex method is more economical.

The flow diagram of the modified simplex method appears in Fig. 2.3.

2.6. The Operating-State Block of Estimation Models

In an estimation model, the operating-state block is the most important one. This is an estimation block because it optimizes the operating state and defines the fuel costs for the optimal operating state.

The operating-state block may be of any one of two types. One optimizes power dispatch, and the other does so for the energy generated by the stations.

The first type of block differs little for conventional analog or digital elements which optimize the operating states of power systems. Yet, there is a difference, and an important one at that — it should be set up to take care of the constraints on power flows between the individual transmission lines. This purpose is usually served by the Lagrangian multiplier method or the gradient method. The fuel costs over a year are determined from a number of typical day total load curves. The selection may include four, eight or twenty-four curves (for a typical week-day and weekend for winter and summer, for each quarter, or for a year). The fuel costs found for each typical day are multiplied by the number of operating days in the year for which they are averages.

The first type of the operating-state block offers the following advantages:

(a) it is possible to assess the effect of network losses on power distribution;

(b) it is possible to trace how the choice of optimum generating units affects the running of the system;

(c) load curves can be obtained for each generating unit as a check on how the desired operating state is maintained;

(d) the constraints on power flows can be conveniently incorporated.

The rationale of the second type of block is to distribute energy by linear programming. It is advantageous in that the algorithm is simple to develop and the computation takes little time to carry out.

Estimation models ordinarily use the first type of block.

In the course of computation, dynamic programming may be utilized in the operating-state block apart from the Lagrangian multiplier and gradient methods.

2.7. Gradient Method

The gradient method is one of the techniques of nonlinear programming. It is applicable to any form of the objective function and the constraint equations, provided the objective function is differentiable and also the dependent variables are differentiable with respect to the independent variables.

Let the objective function be of the form:

$$z=F(x_1, x_2, x_3, \ldots, x_n) \tag{2.38}$$

subject to k constraints ($k<n$) such that

$$\left.\begin{aligned} &W_1(x_1, x_2, \ldots, x_n)=0\\ &W_2(x_1, x_2, \ldots, x_n)=0\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ &W_k(x_1, x_2, \ldots, x_n)=0 \end{aligned}\right\} \tag{2.39}$$

We choose arbitrarily k dependent variables, say $x_1, x_2, \ldots, x_k$, and $(n-k)$ independent variables, say $x_{k+1}, x_{k+2}, \ldots, x_n$. Then, we assign arbitrary values to the independent variables and evaluate all dependent variables and all partial derivatives z of the independent variables:

$$\partial z/\partial x_{k+1}, \qquad \partial z/\partial_{k+2}, \ \ldots, \ \partial z/\partial x_n$$

If these derivatives are all zero or nearly so in value, the solution obtained is an extremal one (which is usually optimal). If the derivatives are nonzero, we should modify all independent variables, using the equations of the gradient method [5], so as to proceed in the direction of the antigradient:

$$\Delta x_j=-\frac{(\partial z/\partial x_j)h}{\sqrt{\sum_{j=k+1}^{n}\left(\frac{\partial z}{\partial x_j}\right)^2}} \tag{2.40}$$

where the size of each step or iteration

$$h=\sqrt{\sum_{j=k+1}^{m}(\Delta x_j)^2} \tag{2.41}$$

is chosen arbitrarily.

We calculate the partial derivatives $\partial z/\partial x_{k+1}, \ldots, \partial z/\partial x_n$ for the changed values of the independent variables, and again modify the independent variables if necessary, and so on.

The procedure is continued until the absolute values of all the listed partial derivatives become less than some sufficiently small value α*. From the values of the independent variables thus obtained, we evaluate the dependent variables and the objective function z by the use of the constraint equations. However, the minimum thus reached may well be local rather than global.

It is only when the objective function is convex that the gradient method discussed leads to a global minimum; but even then it will not happen always. In power-system problems, if the objective function is convex, a global minimum is reached.

The following equations may be used to evaluate the partial derivatives of z with respect to the independent variables:

$$\frac{\partial z}{\partial x_j}=\left(\frac{\partial z}{\partial x_j}\right)'+\sum_{i=1}^{k}\left(\frac{\partial z}{\partial x_i}\right)'\frac{\partial x_i}{\partial x_j} \tag{2.42}$$

$$i=1, 2, \ldots, k; \qquad j=k+1, k+2, \ldots, n$$

where $\partial z/\partial x_j$ = total partial derivative, with all the remaining independent variables assumed constant

$(\partial z/\partial x_j)'$ = partial derivatives, with all remaining variables, both independent and dependent, assumed fixed

Equation (2.42) takes care of the fact that a change in an independent variable x_j entails a change in all dependent variables x_i. The following system of linear equations may be used to evaluate the $\partial x_i/\partial x_j$'s:

$$\frac{\partial W_s}{\partial x_j}=\left(\frac{\partial W_s}{\partial x_j}\right)'+\sum_{i=1}^{k}\left(\frac{\partial W_s}{\partial x_j}\right)\frac{\partial x_i}{\partial x_j}=0 \tag{2.43}$$

$$s=1, 2, \ldots, k \quad (k \text{ equations})$$

The total partial derivative $\partial W_s/\partial x_j$ holds when all the remaining independent variables are held constant, and $(\partial W_s/\partial x_j)'$, when all the other variables are fixed.

Example 2.6. Let there be n electric power stations with active powers $P_1, P_2, \ldots, P_n$. We are to minimize the cost function

$$z=T_1(P_1)+T_2(P_2)+\ldots+T_nP_n \tag{2.44}$$

where $T_i(P_i)$ is the fuel costs as a function of the active power of the ith station.

* This value may be substantially different in different optimization problems.

There is only one constraint:

$$P_1+P_2+\ldots+P_n-P_L-\Delta P=0 \tag{2.45}$$

where the load capacity P_L and the line losses ΔP are deemed constant for simplicity. Let P_n be the dependent variable.

Then

$$\partial z/\partial P_1=\varepsilon_1+\varepsilon_n\partial P_n/\partial P_1$$

$$\partial z/\partial P_2=\varepsilon_2+\varepsilon_n\partial P_n/\partial P_2$$

The partial derivatives $\partial P_n/\partial P_1$, $\partial P_n/\partial P_2$, ... can be evaluated by the equation

$$\frac{\partial W}{\partial P_1}=\left(\frac{\partial W}{\partial P_1}\right)'+\left(\frac{\partial W}{\partial P_n}\right)'\frac{\partial P_n}{\partial P_1}=0$$

or

$$1+1\,\partial P_n/\partial P_1=0$$

whence

$$\partial P_n/\partial P_1=-1$$

Similarly,

$$\partial P_n/\partial P_2=-1, \text{ etc.}$$

Therefore,

$$\partial z/\partial P_1=\varepsilon_1-\varepsilon_n$$

$$\partial_z/\partial P_2=\varepsilon_2-\varepsilon_n, \text{ etc.}$$

By making a guess as to the values of the dependent variables P_1, P_2, ..., P_{n-1}, we then evaluate P_n from the constraint equation. If $\varepsilon_1-\varepsilon_n>0$, and $\varepsilon_2-\varepsilon_n<0$, then changes in P_1 and P_2 can be found by Eq. (2.40):

$$\Delta P_1=-(\varepsilon_1-\varepsilon_n)h\Big/\sqrt{\sum_{j=1}^{n-1}(\varepsilon_j-\varepsilon_n)^2}$$

$$\Delta P_2=-(\varepsilon_2-\varepsilon_n)h\Big/\sqrt{\sum_{j=1}^{n-1}(\varepsilon_j-\varepsilon_n)^2}$$

As is seen, ΔP_1 will be negative and ΔP_2 positive.

The minimization may be terminated when the absolute values of all differences $\varepsilon_j-\varepsilon_n$ are less than the predetermined small value, that is, when the equality

$$\varepsilon_1=\varepsilon_2=\ldots=\varepsilon_n$$

is satisfied for all practical purposes.

There is a difficulty in using the gradient method when some or all constraints are in the form of inequalities. In such a case, we shall be able to proceed in the direction of the antigradient only if all constraints are satis-

fied. Once the boundary of the constraint area is crossed, the movement will be along the projection of the antigradient.

For a global minimum to be reached, both the objective function and the constraint area must be convex.

2.8. Dynamic Programming

Dynamic programming is a form of nonlinear programming, intended to minimize or maximize nonlinear objective functions of many unknowns.

In most of these problems, the constraint equations are linear and the objective function is separable*.

Basically, dynamic programming is a multi-stage process of decision making, in which the function of only one variable is optimized at each stage**. The decision arrived at one stage is stored and used for the next one. The stages in dynamic programming may be the time intervals or the component parts into which the total system has been decomposed for study. The technique can be best illustrated by examples.

Example 2.7. We are to minimize a nonlinear separable function of many unknowns of the form:

$$z=f_1(x_1)+f_2(x_2)+\ldots+f_n(x_n) \tag{2.46}$$

subject to

$$x_1+x_2+\ldots+x_n=X \tag{2.47}$$

Stage one. Evaluate and store $f_1(x_1)$ for x_1 ranging from zero to X.

Stage two. Determine the minimum value of z at zero values of all variables except x_1 and x_2. This is done by minimizing the objective function

$$z_{12}=f_1(x_1)+f_2(x_2)$$

subject to

$$x_1+x_2=X_{12}$$

where X_{12} can take values from zero to X.

Using the constraint equation, the function z_{12} can be defined as

$$z_{12}=f_1(x_1)+f_2(X_{12}-x_1)$$

With X_{12} held constant, z_{12} will depend solely on x_1. The minimum value is

$$z_{12\,\min}=\min\,[f_1(x_1)+f_2(X_{12}-x_1)]$$

The usual methods are not always effective in locating the minimum of a function of a single variable in this class of problems, as it may lie on

* A function of many variables may be called separable if it is equal to the sum of the functions of the constituent variables.

** Optimum distribution between various time instants or entities is ensured over a wide range of likely constraints.

the boundary of the feasible area. Yet, it is always possible to find a minimum value of z_{12} which is a function of $(X_{12}-z_{12\,\min}(X_{12})$, as x_1 takes on values from zero to X_{12}.

In a similar way, we can determine the optimum values for x_1 and x_2, which are likewise functions of X_{12}: $x_{1\,\text{opt}}(X_{12})$ and $x_{2\,\text{opt}}(X_{12})$.

The last three values are stored.

Stage three. Find the minimum of the function

$$z_{13}=f_1(x_1)+f_2(x_2)+f_3(x_3)$$

subject to

$$x_1+x_2+x_3=X_{13}$$

or

$$X_{12}+x_3=X_{13}$$

with X_{13} taking values from zero to X.

Using the constraint equation, we obtain

$$z_{13\,\min}=\min\,[z_{12\,\min}(X_{12})+f_3(X_{13}-X_{12})]$$

With X_{13} held constant, z_{13} will depend solely on X_{12}. As X_{12} takes values from zero to X_{13}, it is always possible to pick $z_{13\,\min}(X_{13})$, and to evaluate $X_{12\,\text{opt}}(X_{13})$, $x_{1\,\text{opt}}(X_{13})$, $x_{2\,\text{opt}}(X_{13})$ and $x_{3\,\text{opt}}(X_{13})$ from the constraint equation.

The succeeding stages are similar to the previous ones.

The last stage gives the values of $z_{\min}(X)$, $x_{1\,\text{opt}}(X)$, $x_{2\,\text{opt}}(X)$, ..., which are obviously the solution of the problem.

As is seen from the foregoing, dynamic programming involves acquisition and storage of excessive data, which is a limitation. The functions stored during each stage are only used at one point. Therefore, dynamic programming is only effective where it can be used repetitively as X takes on a multiplicity of values.

If the objective function z is not separable and the constraint equations are nonlinear (except when the constraint equation contains the product of variables), the problem becomes prohibitively complex for dynamic programming.

In power-system problems, dynamic programming can be used to optimize the operating states of power system, cascaded hydro stations, and expansion schemes.

Example 2.8. We seek to optimize active power dispatch for n electric power stations, assuming constant consumer loads and network losses.

We also assume known the operating costs of the stations as a function of their capacities, $T_i(P_i)$. The total load, including network losses, is P_L.

The objective function has the form

$$z=T_1+T_2+\ldots+T_n=T_1(P_1)+T_2(P_2)+\ldots+T_n(P_n)$$

and is separable. It is to be minimized subject to

$$P_1+P_2+\ldots+P_n=P_L$$

$$P_i \leqslant P_{i\max}$$

Stage one. Note changes in $T(P_1)$ as P_1 takes on values from zero to $P_{1\max}$.

Stage two. Find the minimum

$$T_{12}=T_1(P_1)+T_2(P_{12}-P_1)$$

for $P_1+P_2=P_{12}$, as P_{12} takes on values from 0 to

$$P_{1\max}+P_{2\max}(\text{or to } P_L, \quad \text{if } P_L<P_{1\max}+P_{2\max})$$

As P_1 is varied, evaluate

$$T_{12\,\mathrm{opt}}(P_{12}), \quad P_{1\,\mathrm{opt}}(P_{12}), \quad \text{and} \quad P_{2\,\mathrm{opt}}(P_{12})$$

for each value of P_{12}.

If the functions $T_i(P_i)$ are convex, T_{12} can be minimized, using the equation

$$T_1'(P_1)+T_2'(P_{12}-P_1)=0$$

or

$$\partial T_1/\partial P_1-\partial T_2/\partial P_2=0$$

which corresponds to the equality of specific increments:

$$\varepsilon_1=\varepsilon_2$$

Dynamic programming is also applicable to cases with non-convex functions. In such cases, P_1 is modified for each fixed value of P_{12}, while the remaining stages are the same. The last stage yields $T_{\mathrm{opt}}(P_L)$, $P_{1\,\mathrm{opt}}(P_L)$, $P_{2\,\mathrm{opt}}(P_L)$, which constitutes the solution of the problem.

As is seen from the foregoing, every next stage takes up one more station for consideration.

Example 2.9. We seek to optimize economic dispatch for a system containing n thermal power stations and one hydro with the flow of water constrained to B per day.

We assume known the relation $T(P_\Sigma)$ for all thermal stations, where P_Σ is the sum of their active powers (see Example 2.8). Also, we know the total load on the system over twenty-four hours, $P_{L1}, P_{L2}, \ldots, P_{L24}$.

For the hydro, we know its power is a function of hourly flow rate of water for the ith hour, that is, $P_{hi}=f(B_i)$.

We shall solve the problem by use of dynamic programming.

Stage one: Note the behaviour of the relation $P_{h1}(B_1)$, where the hourly flow rate of water, B_1, takes on values from zero to B, or as the hydrostation power, P_{h1}, takes on values from zero to $P_{h\max}$.

The costs as a function of B_1 for the thermal stations during the first hour are

$$T_1[P_{L1}-P_{h1}(B_1)]=T_1(B_1)$$

This relation is stored.

Stage two: Minimize the costs for the thermal stations over two hours:

$$T_{12}=T_1(P_1)+T_2(P_2)=T_1(P_1)+T_2(P_{L2}-P_{h2})$$

subject to $B_1+B_2=B_{12}$, where B_{12} can take on values from zero to B (or constrained by the maximum capacity of the hydro station).

Using the constraint equations, we obtain

$$T_{12}=T_1[P_{L1}-P_{h1}(B_1)]+T_2[P_{L2}-P_{h2}(B_{12}-B_1)]$$

While holding B_{12} constant and varying B_1, it is possible to find optimal values for $B_{1\,\mathrm{opt}}(B_{12})$, $B_{2\,\mathrm{opt}}(B_{12})$, $P_{1\,\mathrm{opt}}(B_{12})$, $P_{2\,\mathrm{opt}}(B_{12})$ and $T_{12\,\mathrm{opt}}(B_{12})$.

Stage three: Minimize the costs over three hours:

$$T_{13}=T_1(P_1)+T_2(P_2)+T_3(P_3)$$

subject to

$$B_{13}=B_1+B_2+B_3=B_{12}+B_3$$

where B_{13} is allowed to take values from zero to B or is constrained by the maximum capacity of the hydrostation. Using the constraint equations, we obtain

$$T_{13}=T_{12\,\mathrm{opt}}(B_{12})+T_3[P_{L3}-P_{h1}(B_{13}-B_{12})]$$

While holding B_{13} constant and allowing B_{12} to take different values, we evaluate $B_{1\,\mathrm{opt}}(B_{13})$, $B_{2\,\mathrm{opt}}(B_{13})$, ..., $P_{1\,\mathrm{opt}}(B_{13})$, $P_{2\,\mathrm{opt}}(B_{13})$, ..., $T_{13\,\mathrm{opt}}(B_{13})$.

The succeeding stages are similar to the preceding ones. The last stage yields $B_{1\,\mathrm{opt}}(B)$, $B_{2\,\mathrm{opt}}(B)$, ..., $P_{1\,\mathrm{opt}}(B)$, $P_{2\,\mathrm{opt}}(B)$, ..., $T_{\mathrm{opt}}(B)$, which constitutes the solution of the problem.

In the above problem, the multi-stage process of decision-making was arranged as a time series, and one hour was added at each step.

2.9. Network Block

The network block serves to determine the discounted expansion and operating costs of the main transit networks. Distribution networks are not included usually in consideration, because their costs are almost independent of station siting. The network block is likewise an optimization model, as it is intended to determine the costs of an optimal network. It this case, the objective function has terms for the construction costs of new networks, the operating costs and the network loss costs. The final costs of the network after optimization do not include the network losses (so as not to take them into account twice, because the operating-state block will lump them with the fuel costs). If the cost function is convex, the network block may well use the gradient method; if the function is non-convex, resort will have to be made to the branch-and-bound method.

Assuming that the cost function is linear, approximate computation, notably in the case of linear models, can proceed by use of linear programming as adapted to transportation problems. Estimation models can also use solution by partial enumeration, which does not however lead to a global minimum in all cases.

The operating-state and network blocks of an estimation model are related as follows. The network block should lead the operating-state block in time. Because the operating-state block takes account of network losses, it should be fed data on the state of the network. However, the state of the network is determined for a particular level of operation for which an optimum power dispatch between the stations must be known. Therefore, the procedure starts by guessing the power dispatch that should be put into the network block. After the operating-state block performs the first iteration, a correction is applied to the network block, if necessary, and the next iteration step is carried out. In other words, there must be feedback between the operating-state and network blocks.

It is also important that only one network configuration be chosen, although many operating states can occur over a year. This is why a network is usually assessed in connection with several operating states. If the optimum network configurations thus obtained differ markedly for different operating states, the final decision is arrived by testing each alternative network configuration in all operating states, and the annual fuel costs must be found by considering the constraints on power flow in some network elements in certain operating states. By adding together the network and fuel costs, it is possible to pick an optimum network configuration and to determine the total optimum costs for these items.

In principle, the cost of loss to consumers due to interruption of service should be included in the objective function of the network block. However, this would make the network block extremely complicated. Therefore, a compromise solution is adopted, which consists in that a separate cost-of-loss block is set up, so that the need for network reinforcement can be assessed in order to enhance its security (provision of double instead of single circuits, use of closed ring networks, etc.).

The dynamics of the discounted network construction costs can be taken into account, at least approximately, by one of two methods.

In the first method, after the static network problems are solved for every year of the period, the next step is to determine which network elements are stable and unstable (that is, optimal at the beginning of the period and nonoptimal thereafter). Two alternatives are considered for unstable elements: one with all unstable elements included in a complete graph, and the other with these elements eliminated from the complete graph. In this way, a decision can be made whether to build such elements or not.

In the second method, one solves the static problem associated with the last year of the period under consideration, and the network thus developed is assumed to be the complete graph for the preceding years.

2.10. Transportation Problems

Transportation problems comprise a form of linear programming. Their distinction is that the system of constraint equations is very simple, the coefficients of all variables being 1 or 0. In view of this, transportation problems can be solved by use of simpler algorithms than the simplex algorithm.

To begin with, we shall examine the *classical* transportation problem and its algorithm.

In this problem, we have m sources of a commodity with fixed inventories a_i, where $i=1, 2, \ldots, m$; and n consumers of this product with fixed demands b_j where $j=1, 2, \ldots, n$. As applied to the power industry, these may be m electric power stations and n load centres.

The commodity is shipped (power is transmitted) only from suppliers to consumers, and the cost of shipment (transmission) is proportional to the quantity of commodity shipped.

Then the objective function takes the form

$$z=c_{11}x_{11}+c_{12}x_{12}+\ldots+c_{1n}x_{1n}+c_{21}x_{21}+c_{22}x_{22}+\ldots+c_{2n}x_{2n} + \ldots+c_{m1}x_{m1}+c_{m2}x_{m2}+\ldots+c_{mn}x_{mn} \qquad (2.48)$$

where x_{ij} = quantity of commodity shipped from the ith supplier to the jth consumer

c_{ij} = transportation cost per unit of commodity for this shipment

The constraint equations take the form of nodal balance equations:

$$\sum_{j=1}^{n} x_{ij}=a_i, \qquad \sum_{i=1}^{m} x_{ij}=b_j, \qquad \sum_{i=1}^{m} a_i=\sum_{j=1}^{n} b_j \qquad (2.49)$$

The last equation describes the perfect balance of all supplies and demands. In the case of imbalance, the number of basic variables is $m+n$, and in the case of balance it is $m+n-1$, because then one of the node balance equations may be discarded.

To solve the problem, the constraint equations are written in the form of a transportation matrix where the rows correspond to the balance equations at the sources and the columns to the balance equations at the consumers. Besides, the transportation costs (per unit) are entered in the lower right-hand corner of each cell.

For three suppliers (1, 2 and 3) and three consumers (1, 2 and 3) the transportation matrix has the form shown on page 141.

The right-hand sides of the balance equations are written in the right-hand column for the suppliers and in the lower row for the consumers. In this way, the matrix gives a compact representation for the following constraint equations:

$$x_{11}+x_{12}+x_{13}=a_1 \qquad x_{11}+x_{21}+x_{31}=b_1$$
$$x_{21}+x_{22}+x_{23}=a_2 \qquad x_{12}+x_{22}+x_{32}=b_2$$
$$x_{31}+x_{32}+x_{33}=a_3 \qquad x_{13}+x_{23}+x_{33}=b_3$$

	1	2	3	a_i	
1	x_{11} c_{11}	x_{12} c_{12}	x_{13} c_{13}	a_1	u_1
2	x_{21} c_{21}	x_{22} c_{22}	x_{23} c_{23}	a_2	u_2
3	x_{31} c_{31}	x_{32} c_{32}	x_{33} c_{33}	a_3	u_3
b_j	b_1	b_2	b_3		
	v_2	v_2	v_3		

The terms u_1, u_2, u_3 and v_1, v_2, v_3 are called *simplex factors.* No equation is written for the objective function.

Strictly speaking, the equations are not in canonical form. At perfect balance, there will be five basic variables (out of the total number of nine) and four non-basic variables.

In solving a transportation problem, the first step is to guess an arbitrary initial basic solution. This solution should contain five basic shipments and satisfy all the constraint equations. Although the guess may be any, computation time can be saved if this is done in accordance with the following rules:

(1) choose in the transportation matrix the cell with the least value of c_{ij};

(2) as a basic term, use $x_{ij}=\min(a_i, b_j)$, that is, the lowest term corresponding to perfect balance between the ith supplier and the jth consumer. This will satisfy the balance equations of the row or column containing x_{ij}. Write zeros in the cells assigned to the remaining unknowns in the equation satisfied. If, for example, $b_j<a_i$, then $x_{ij}=b_j$, $x_{1j}=x_{2j}=\ldots=0$ (except x_{ij}). Then b_j is replaced by zero, and a_i by (a_i-b_j);

(3) choose again the cell holding the lowest value of c_{ij} from among the cells not yet holding numerical values for the shipments, and go through steps 2 and 3 until all shipments are evaluated (that is, until all non-basic variables are reduced to zero).

After the initial basic solution has thus been obtained, it is optimized by use of the simplex coefficients u_i and v_j. The simplex coefficients have the property that if all horizontal constraint equations be multiplied respectively by u_1, u_2, and u_3, and all vertical constraint equations respectively by v_1, v_2 and v_3, all the equations be added together, and the sum be subtracted from the objective function, the new values of the coefficients c_{ij} for all basic unknowns will be zero, that is the basic unknowns will be eliminated from the objective function. The coefficients c_{ij} for the non-basic variables will take new values. Designating the new values of the

coefficients c_{ij} obtained as a result of this transformation as $\bar{c}_{ij}$ (with an overscribed bar), we can trace a certain relationship between the $\bar{c}_{ij}$'s, the c_{ij}'s, and the simplex coefficients.

As a proof, after the transformation we have

$$z=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}-\sum_{i=1}^{m}\left[u_i\left(\sum_{j=1}^{n}x_{ij}\right)-u_ia_i\right]-\sum_{j=1}^{n}\left[v_j\sum_{i=1}^{m}x_{ij}-v_jb_j\right]$$

or

$$z-\sum_{i=1}^{m}a_iu_i-\sum_{j=1}^{n}b_jv_j=\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij}-u_i-v_j)x_{ij}=\sum_{i=1}^{m}\sum_{j=1}^{n}\bar{c}_{ij}x_{ij} \tag{2.50}$$

where $\bar{c}_{ij}=c_{ij}-u_i-v_j$.

If the simplex coefficients are chosen such that for the basic unknowns they are equal to zero, then

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\bar{c}_{ij}x_{ij}=0$$

because all non-basic unknowns are equal to zero. Thus,

$$z=\sum_{i=1}^{m}a_iu_i+\sum_{j=1}^{n}b_jv_j \tag{2.51}$$

The simplex coefficients of all basic unknowns can be evaluated by use of equations of the form

$$\bar{c}_{ij}=c_{ij}-u_i-v_j=0 \tag{2.52}$$

or

$$u_i+v_j=c_{ij} \tag{2.53}$$

There are $(m+n-1)$ such equations, while the number of coefficients is $(m+n)$. Therefore, one of the coefficients may be set equal to zero. This amounts to removal of one of the constraint equations at perfect balance of the commodity. In this way, all simplex coefficients can be evaluated.

Once all simplex coefficients are known, we can determine the new values of the coefficients $\bar{c}_{ij}$ for all non-basic unknowns. If all of the new coefficients are non-negative, the initial basic solution is optimal, because the introduction of any non-basic variables into the basic cannot minimize the objective function any more. If at least one of the coefficients $\bar{c}_{ij}$ is negative, the introduction of the shipping unknowns x_{ij} into the basis will minimize the objective function. To speed up the optimization procedure, the shipping unknown, x_{st}, to be introduced into the basis can be chosen from the condition

$$\bar{c}_{st}=\min(\bar{c}_{ij}), \qquad \bar{c}_{st}<0 \tag{2.54}$$

The unknown to be eliminated from the basis is chosen as follows. Let $x_{st}=\theta$, where θ is the term introduced into the shipping basis. Then, in order to preserve nodal balance, a basic unknown in the s-row and the t-column should be replaced by θ. Next, to make up for the slack at other

nodes, the corrections $+\theta$ are applied to the respective basic unknowns. As a rule, balance is restored when application of corrections $\pm\theta$ produces a closed polygon. This polygon corresponds to the formation of a closed loop by addition of a branch. The corrections $+\theta$ do not, and the corrections $-\theta$ do limit the value of θ. When the value of θ in the cell where the correction $-\theta$ is applied becomes equal to the original shipping value in that cell, the resultant shipment reduces to zero and is eliminated from the basic variables. The maximum value of θ will then be equal to the minimum value of the basic variables in the cells where the correction $-\theta$ is applied. Obviously, the reduction in z will be $|\bar{c}_{st}|\theta$.

After θ is evaluated and the basic variables are adjusted, the next step is to find the changed values of the simplex coefficients u_i and v_j, and the new values of the $\bar{c}_{ij}$'s. If at least one of them is negative, the unknown to be introduced into the basis is identified, and so on. The optimization procedure is complete when all the c_{ij}'s evaluated for the non-basic variables are non-negative.

Example 2.10. There are three electric power stations with the following capacities (in MW):

$$a_1=25, \qquad a_2=25, \qquad a_3=50$$

and four consumers with demands (MW):

$$b_1=15, \qquad b_2=20, \qquad b_3=30, \qquad b_4=35$$

As is seen,

$$\sum a_i=\sum b_j=100$$

The transmission cost per unit factors are:

$$c_{11}=10, \quad c_{12}=5, \quad c_{13}=6, \quad c_{14}=7, \quad c_{21}=8, \quad c_{22}=2,$$
$$c_{23}=7, \quad c_{24}=6, \quad c_{31}=9, \quad c_{32}=3, \quad c_{33}=4, \quad c_{34}=8$$

	1	2	3	4	a_i
1	0_{10}	0_5	0_6	25_7	25
2	0_8	20_2	0_7	5_5	25
3	15_9	0_3	30_4	5_8	50
b_j	15	20	30	35	

Find the initial solution.

The minimum c_{ij} is $c_{22}=2$. Enter the least of the two a_i and b_j in cell 2-2, that is, out of the values 25 and 20, choose $x_{22}=20$. Now cell 2-2 holds the value 20, while cells 1-2 and 1-3 hold zeros. Next, delete $b_2=20$ and enter $a_j=25$ with $5=(25-20)$. Now column 2 is filled. Next, the minimum

c_{ij} is c_{33}. Enter $x_{33}=\min(50, 30)=30$, that is, $x_{33}=30$, and write zeros in cells 1-3 and 2-3. Then delete $b_3=30$, and replace $a_3=50$ with 20. Next, $c_{ij\,\min}=c_{24}$. Introduce $x_{24}=\min(5, 35)$, that is $x_{24}=5$. Place 5 in cell 2-4, and write zero in cell 2-1. Delete $a_2=5$, and replace $b_4=35$ with 30. Now $c_{ij\,\min}=c_{14}$, and $x_{14}=\min(30, 25)=25$. Place $x_{14}=25$ in cell 1-4, and write zero in cell 1-1. Delete $a_1=25$ and replace $b_4=30$ with 5. Next $c_{ij\,\min}=c_{34}$. Introduce $x_{34}=5$. Delete $b_4=5$ and write 15 instead of $a_3=20$. The only cell remaining is cell 3-1. Place $x_{31}=15$ in it.

The basic solution thus obtained is: $x_{14}=25$, $x_{22}=20$, $x_{24}=5$, $x_{31}=15$, $x_{33}=30$, $x_{34}=5$. The non-basic variables are $x_{11}=x_{12}=x_{13}=x_{21}=x_{23}=x_{32}=0$.

The simplex coefficients are evaluated by the following equations:

$$u_1+v_4=7$$
$$u_2+v_2=2$$
$$u_2+v_4=6$$
$$u_3+v_1=9$$
$$u_3+v_3=4$$
$$u_3+v_4=8$$

We let $v_4=0$ (see above). Then,

$$u_1=7, \quad u_2=6, \quad u_3=8, \quad v_1=1, \quad v_2=-4, \quad v_3=-4, \quad v_4=0$$

Now we evaluate z [see Eq. (2.42)]:

$$z=20\times2+5\times6+25\times7+5\times3+30\times4+15\times9=540$$

The same result is obtained by use of Eq. (2.45):

$$z=25\times7+25\times6+50\times8+15\times1-20\times4-30\times4=540$$

Evaluate the $\bar{c}_{ij}$'s for the non-basic variables:

$$\bar{c}_{11}=10-7-1=2>0, \qquad \bar{c}_{21}=8-6-1=1>0$$
$$\bar{c}_{12}=5-7+4=2>0, \qquad \bar{c}_{23}=7-6+4=5>0$$
$$\bar{c}_{13}=6-7+4=3>0, \qquad \bar{c}_{32}=3-8+4=-1<0$$

To minimize z, we introduce $x_{32}=\theta$. As a result, we get

	1	2	3	4	a_i
1				25	25
2		$20-\theta$		$5+\theta$	25
3	15	$+\theta$	30	$5-\theta$	50
b_j	15	20	30	35	

To maintain balance, we enter $-\theta$ in cells 2-2 and 3-4, and $+\theta$ in cell 2-4. This will ensure balance at all nodes. The maximum value of $\theta = \min(20, 5) = 5$. That is, $\theta = 5$.

The new basic solution is, $x_{14}=25$, $x_{22}=15$, $x_{24}=10$, $x_{31}=15$, $x_{32}=5$, $x_{33}=30$.

The non-basic unknowns are: x_{11}, x_{12}, x_{13}, x_{21}, x_{23}, x_{34}.

Evaluate the new simplex-coefficients by the equations:

$$u_1+v_4=7$$
$$u_1+v_2=2$$
$$u_2+v_4=6$$
$$u_3+v_1=9$$
$$u_3+v_2=3$$
$$u_3+v_3=4$$

Setting $v_4=0$, we obtain: $u_1=7$, $u_2=6$, $u_3=7$, $v_1=2$, $v_2=-4$, $v_3=-3$, $v_4=0$.

Evaluate $\bar{c}_{ij}$ for the non-basic unknowns:

$$\bar{c}_{11}=10-7-2=1>0$$
$$\bar{c}_{12}=5-7+4=2>0$$
$$\bar{c}_{13}=6-7+3=2>0$$
$$\bar{c}_{21}=8-6-2=0$$
$$\bar{c}_{23}=7-6+3=4>0$$
$$\bar{c}_{24}=8-7+0=1>0$$

that is, the solution obtained is optimal.

The value of the objective function is

$$z=25\times 7+25\times 6+50\times 7+2\times 15-4\times 20-3\times 30=535$$

The change in z is $\theta=|c_{32}|=5(-1)=-5$, which is true.

The classical transportation-problem approach may be used also where no perfect balance exists, for example, when a system has some reserve capacity, that is, when

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$$

In such a case, however, the constraints on the load centres (nodes) will have to be written as inequalities, that is, perfect balance equation will have to be abandoned.

For three stations ($i=1, 2, 3$) and for three load centres ($j=1, 2, 3$) the transportation matrix has the form

x_{10}	x_{11}	x_{12}	x_{13}	a_1	u_1
x_{20}	x_{21}	x_{22}	x_{23}	a_2	u_2
x_{30}	x_{31}	x_{32}	x_{33}	a_3	u_3
$\Sigma a_i - \Sigma b_j$	b_1	b_2	b_3		
	v_1	v_2	v_3		

where x_{i0} is the surplus capacity of (slack for) the ith station and the coefficients c_{10}, c_{20} and c_{30} are equal to zero. The number of basic unknowns is $(m+n)$.

The solution of the problem should yield optimum station capacities and optimum blocks of power transmitted.

The transportation technique may be modified or extended, by adding to the coefficients c_{ij} the unit generation cost c_i at the ith station. Then the new value will be

$$c_{ij}^{(0)} = c_{ij} + c_i \tag{2.55}$$

The transportation technique may be used in power-system analysis to give a rough approximation to the optimum system structure. The generation costs per unit c_i are equal to the discounted investment, depreciation, operating and fuel costs per unit*.

The problem is solved, assuming that the total station capacity exceeds the total demand or load.

The solution of such a problem defines optimum station capacities and optimum network configurations. The solution is approximate because we ignore the fact that the station construction, fuel and network costs are nonlinear. Because of this, the modified, or extended, transportation technique is applicable only to cases where it is desired to obtain a rough outline of the basic system structure.

The transportation technique is applicable to one more class of transportation problems involving trans-shipments. In contrast to the classical transportation problem, shipments are permitted from one point of origin to another and also from one point of destination to another. All nodes, whether origins or destinations, are numbered consecutively, $j=1, 2, 3, \ldots, n$. A specific feature of this type of problem is that a trans-shipped commodity may turn up at some nodes. This may be so at a point of origin if shipment is in the direction of that point, or at a point of destination, if shipment is from a point of destination.

* No fuel costs are entered in the first column giving the surplus capacity (reserve or slack).

The objective function has the form

$$z=\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}c_{ij}x_{ij} \tag{2.56}$$

For the jth point of origin, the constraint equation has the form

$$\sum_{k}x_{jk}-x_{jj}=a_j \tag{2.57}$$

where j = index of a given node
k = index of other nodes to which shipment is directed from the jth node
x_{jj} = trans-shipped commodity
a_j = commodity inventory at the jth node

If shipments are from the jth node to all other points, then

$$x_{jj}=0$$

and

$$\sum_{k}x_{jk}=a_j \tag{2.58}$$

If there are shipments from other nodes, i, to point j, their sum gives the trans-shipped commodity

$$x_{jj}=\sum_{i}x_{ij} \tag{2.59}$$

where i designates the nodes from which shipments are directed to point j. In such a case, the constraint equation takes the form of Eq. (2.57).

For the points of destination the constraint equation is

$$\sum x_{ij}-x_{jj}=b_j \tag{2.60}$$

where b_j = demand at the jth node
i = designates the nodes from which the commodity is shipped to the jth point

If shipments are routed from all remaining points to point j, then

$$x_{jj}=0$$

and

$$\sum_{i}x_{ij}=b_j \tag{2.61}$$

In the case of shipments from the jth point of destination to other nodes, k, their sum gives the trans-shipped commodity

$$x_{jj}=\sum_{k}x_{jk} \tag{2.62}$$

and the respective constraint equation takes the form of Eq. (2.60).

If we assume that the trans-shipped commodities enter in the objective function with zero coefficients c_{jj}, then the constraint equations may be arranged in the following tableau:

	1	2	3	...	n	a_j	
1	$-x_{11}$	x_{12}	x_{13}	...	x_{1n}	a_1	$-\pi_1$
2	x_{21}	$-x_{22}$	x_{23}	...	x_{2n}		$-\pi_2$
3	x_{31}	x_{32}	$-x_{33}$	...	x_{3n}	a_3	$-\pi_3$
...	...	...	...	...	...	...	
n	x_{n1}	x_{n2}	x_{n3}	...	$-x_{nn}$		$-\pi_n$
b_i		b_2		...			
	π_1	π_2	π_3		π_n		

In the tableau, the trans-shipped commodities are entered with the '—' sign, in accordance with Eqs. (2.57), (2.59), (2.60) and (2.62). The a_j column represents the capacity of the points of origin, and the b_i row the requirements of the points of destination. In the above tableau, cells 1 and 3 are origins, and cell 2 is a destination. The rows correspond to Eqs. (2.57), written for origins and to Eqs. (2.60) written for destinations. The columns correspond to Eqs. (2.62) for destinations and to Eqs. (2.59) for origins. The objective function has the form

$$z=\sum_{i=1}\sum_{j=1}c_{ij}x_{ij}$$

where, as distinct from Eq. (2.56), i may be equal to j (when c_{jj} is zero). Each cell holds one of the coefficients c_{ij}. As a rule, $c_{ij}=c_{ji}$. It is to be noted that the simplex coefficients for the equally numbered rows and columns have the same absolute value but differ in sign, so that $c_{jj}=0$. These coefficients are called *node potentials*. For the basic unknowns

$$c_{ij}=\pi_j-\pi_i \tag{2.63}$$

If a trans-shipped commodity, x_{jj}, is included in the basis, then

$$c_{jj}=\pi_j-\pi_j=0 \tag{2.64}$$

Illegitimate shipments may be shaded in the tableau. In this type of problem, the optimization proceeds in the same manner as in the classical transportation problem. The initial feasible solution may be either guessed or found by the minimum values of c_{ij} (see above). Then, subject to (2.63) for the basic unknowns, the potentials π_j are evaluated (one of them

may be set equal to zero) and the $\bar{c}_{ij}$'s are found for all non-basic unknowns by the equation

$$\bar{c}_{ij}=c_{ij}-(-\pi_i)-\pi_j)=c_{ij}+\pi_i-\pi_j \tag{2.65}$$

If all $\bar{c}_{ij}$ for the non-basic unknowns are non-negative, the solution obtained is optimal. If some $\bar{c}_{ij}<0$, a shipment, x_{st}, is introduced into the basis, subject to the condition

$$\bar{c}_{st}=\min(\bar{c}_{ij}) \tag{2.66}$$

such that

$$\bar{c}_{st}<0$$

Then $x_{st}=\theta$ is evaluated on the condition that the correction applied to any basic unknown will reduce it to zero, that is, eliminate it from the basis. After that, the potentials are evaluated by Eq. (2.63) and the $\bar{c}_{ij}$'s are re-evaluated for the new basic variables. The optimization procedure is complete when all $\bar{c}_{ij}$ for the non-basic variables are non-negative.

Example 2.11. This is an example of a power-system problem involving transit power transmission.

Let there be electric power stations with capacities (MW) $a_1=100$ and $a_5=200$, and with loads $b_2=160$, $b_3=80$, and $b_4=60$. Let also the transmissions x_{15}, x_{24}, x_{42} and x_{51} be illegitimate.

In the tableau below, the coefficients c_{ij} are written in the corners of the cells; the cells representing the illegitimate transmissions are shaded.

	1	2	3	4	5	a_j	
1	0	5	6	100_2		100	$-\pi_1$
2	5	-40_0	40_3		3		$-\pi_2$
3	6	3	0	4	4		$-\pi_3$
4	2		40_4	-40_0	7		$-\pi_4$
5		200_3	4	7	0	200	$-\pi_5$
b_j		160	80	60			
	π_1	π_2	π_3	π_4	π_5		

To begin with, we guess the following initial feasible solution: $x_{14}=100$, $x_{43}=40$, $x_{23}=40$, and $x_{52}=200$. To keep balance between the rows, we introduce $x_{22}=-40$ and $x_{44}=-40$. In accordance with Eqs. (2.56) and

(2.60), the trans-shipped commodity at point *2* is $100-60=40$, and at point *4* it is $200-60=140$. Evaluate the node potentials by the equations

$$\pi_4-\pi_1=2$$
$$\pi_3-\pi_2=3$$
$$\pi_3-\pi_4=4$$
$$\pi_2-\pi_5=3$$

Setting $\pi_5=0$, we get

$$\pi_1=0, \quad \pi_2=3, \quad \pi_3=6, \quad \pi_4=2$$

Evaluate the $\bar{c}_{ij}$'s for all non-basic variables:

$$\bar{c}_{12}=5+0-3=\ 2>0 \qquad \bar{c}_{13}=6+0-6=\ 0$$
$$\bar{c}_{21}=5+3-0=\ 8>0 \qquad \bar{c}_{25}=3+3-0=\ 6>0$$
$$\bar{c}_{31}=6+6-0=12>0 \qquad \bar{c}_{32}=3+6-3=\ 6>0$$
$$\bar{c}_{34}=4+6-2=\ 8>0 \qquad \bar{c}_{35}=4+6-0=10>0$$
$$\bar{c}_{41}=2+2-0=\ 4>0 \qquad \bar{c}_{45}=7+2-0=\ 9>0$$
$$\bar{c}_{53}=4+0-6=-2<0 \qquad \bar{c}_{54}=7+0-2=\ 5>0$$

Introduce $\bar{x}_{53}$ into the basis, because $\bar{c}_{53}<0$. As is seen from the tableau, x_{23} is eliminated from the basis, and $\theta=40$. The new feasible solution is: $x_{14}=100$, $x_{43}=40$, $x_{44}=-40$, $x_{52}=160$, and $x_{53}=40$.

The new values of node potentials can be found from the equations:

$$\pi_4-\pi_1=2, \quad \pi_3-\pi_4=4, \quad \pi_2-\pi_5=3, \quad \pi_3-\pi_5=4$$

Setting $\pi_5=0$, we get

$$\pi_1=-2, \quad \pi_2=3, \quad \pi_3=4, \quad \pi_4=0$$

Evaluate the $\bar{c}_{ij}$'s for all non-basic variables:

$$\bar{c}_{12}=5-2-3=0 \qquad \bar{c}_{13}=6-2-4=0$$
$$\bar{c}_{21}=5+3+2=10>0 \qquad \bar{c}_{23}=3+3-4=2>0$$
$$\bar{c}_{25}=3+3-0=\ 6>0$$
$$\bar{c}_{31}=6+4+2=12>0 \qquad \bar{c}_{32}=3+4-3=4>0$$
$$\bar{c}_{34}=4+4-0=\ 8>0 \qquad \bar{c}_{35}=4+4-0=8>0$$
$$\bar{c}_{41}=2+0+2=\ 4>0 \qquad \bar{c}_{42}=7+0-0=7>0$$
$$\bar{c}_{54}=7+0-0=\ 7>0$$

This solution is optimal, because all $\bar{c}_{ij}\geqslant 0$.

For the initial feasible solution

$$z=100\times 2+40\times 3+40\times 4+200\times 3=1080$$

For the optimal solution:

$$z=100\times 2+40\times 4+160\times 3+40\times 4=1000<1080$$

The change in z is

$$|c_{st}|\theta=2\times 40=80$$

The technique used to solve the transportation problem involving trans-shipments may be applied to the network transportation problem for which the shipping cost is nonlinear:

$$C_{ij}=\begin{cases}0 & \text{if } x_{ij}=0\\ a_{ij}+b_{ij}x_{ij} & \text{if } x_{ij}\neq 0\end{cases} \tag{2.67}$$

where a_{ij} is the fixed component in the cost of shipping from i to j.

In this case,

$$c_{ij}=a_{ij}/x_{ij}+b_{ij}$$

The objective function has the form

$$z=\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}a_{ij}+\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}b_{ij}x_{ij} \tag{2.68}$$

The resultant optimal network has $(n-1)$ branches and is a tree. This is so because if a network has l branches such that $l>n-1$, then applying the technique used in the transportation problem with trans-shipments to

$$z=\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}b_{ij}x_{ij} \tag{2.69}$$

we obtain an optimum basic solution with $(n-1)$ basic equations. As a consequence, shipments along $l-(n-1)$ branches will be zero. In view of condition (2.67), however, in these branches $a_{ij}\neq 0$, that is, it would give no advantage to incorporate these branches in the network. Thus, the optimum solution corresponds to a tree.

The initial tree can be chosen arbitrarily. In this tree, the shipments will be defined uniquely. Sometimes, it will be a good plan to use a tree corresponding to $\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\equiv\min$. Next, we evaluate the potentials π for this tree, and also

$$\bar{b}_{ij}=b_{ij}+\pi_i-\pi_j \tag{2.70}$$

for all links (that is, branches which are not part of the tree).

For each link ij it is possible to determine the branches rs obtained from the basic solution by inspection of the loop formed. The shipment is assumed to be zero in value because θ is a circulating flow around the loop. At $\theta=\theta_{\min}$, the direction of shipping in the other branches of the loop does not change. The introduction of the link ij with flow θ and the removal of the branch rs from the network causes the shipping cost to change by

$$\Delta C=-a_{rs}+a_{ij}+\theta b_{ij}$$

If $\Delta C<0$, the introduction of the link ij to the tree reduces z. The link to be introduced should be that which secures the greatest reduction in z. Then the potentials π and b_{ij} are again evaluated for the links of the new tree, and so on. However, this technique leads to a local, and not a global, extremum [2].

2.11. Branch and Bound Method

The branch and bound method is a technique of mathematical programming which leads to a global extremum (say, a minimum) when applied to nonlinear multi-extremal problems. Although primarily intended to handle combinatorial extremal problems, the method may well handle optimization problems with both continuous and discrete unknowns.

The branch and bound method reduces basically to the following. A decision tree is constructed, which has an initial node associated with the set Q of all feasible decisions or solutions. From this node, the set branches into subsets, Q_1 and Q_2, such that

$$Q_1+Q_2=Q$$

Each of the two nodes, Q_1 and Q_2, can serve as a start for further branching. The final node represents a single decision or solution. The sum of the subsets of all hanging nodes in the course of branching is equal to Q. At each node, the lower bound, LB, is put on the objective function, OF, and the precise value is found for it at the final node. The name of the method, branch and bound, arises from the fact that the tree of alternatives is composed of branches and that the lower bound is put on each of its nodes.

If branching produces a hanging node for which $LB\geqslant OF_0$, where OF_0 is the optimal value of the objective function at one of the final nodes obtained previously, the branching need not be continued from the hanging node, and the entire subset of decisions associated with that node, however large, may be discarded, because it cannot contain solutions better than OF_0.

If, at a particular stage of branching, $LB\geqslant OF_0$ for all hanging nodes, the minimization procedure is complete. The OF_0 now corresponds to a global optimal solution. In fact, this is an enumeration of all solutions taken in subsets rather than individually, but it secures finding a global extremum all the same.

Various branching algorithms and various methods for deriving the lower bound exist, and their selection depends on the problem at hand. As the accuracy of the lower bound is improved, fewer hanging nodes need be examined. However, it takes more effort to compute the lower bound.

Example 2.12. An optimum electric network is to be chosen at the costs defined (leaving out the existing networks) as

$$C_{ij}=\begin{cases} 0 & \text{for } P_{ij}=0 \\ a_{ij}+b_{ij}P_{ij} & \text{for } P_{ij}>0 \end{cases} \tag{2.71}$$

where C_{ij} = discounted cost per network element (line) ij from point i to point j
P_{ij} = active power flow in branch ij
a_{ij}, b_{ij} = constants specified in advance

The objective function may be written as

$$z=\sum_{ij} C_{ij}(P_{ij})=\sum_{ij} a_{ij}+\sum_{ij} b_{ij}P_{ij} \tag{2.72}$$

where the summation is taken over all branches of the complete graph. The constraint equation is

$$\sum_{j} P_{ij}-\sum_{j} P_{ji}=P_j \tag{2.73}$$

As has been shown in Sec. 2.9, an optimal network with such costs is a tree, that is, it contains no loops (meshes).

The following branching procedure will be adopted. From any hanging node, branching will proceed in two directions. One direction, say to the right, will represent the participation of a branch, rs, in the tree, and the other, to the left, will represent the suppression of the same branch, designated as $\overline{rs}$. Thus, the set Q_i is partitioned into two subsets, Q_{rs} (branch rs being part of the tree) and $Q_{\overline{rs}}$ (branch $\overline{rs}$ being forbidden) (Fig. 2.4).

For node i, the lower bound, LB_i, is

$$LB_i=\min \sum_{Q_i} a_{ij}+\min \sum_{Q_i} b_{ij}P_{ij}\leqslant\left[\sum_{Q_i}(a_{ij}+b_{ij}P_{ij})\right] \tag{2.74}$$

The condition $\min \sum_{Q_i} a_{ij}$ defines a tree for which the sum of all a_{ij} within the set Q_i is a minimum. Let this tree be designated as D_1. The condition $\min \sum_{Q_i} b_{ij}P_{ij}$ defines a tree for which the linear transportation technique within the set Q_i yields an optimal solution. Let this tree be designated as D_2.

If both trees, D_1 and D_2, are coincident, the coincidence tree corresponds to an optimal solution within the set Q_i. If both are coincident at the initial node, the coincidence tree corresponds to the solution of the problem.

With the branching procedure defined above, a tree of alternative decisions will have the form shown in Fig. 2.5, where O is the initial node,

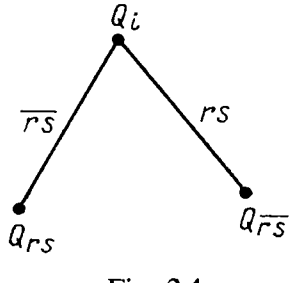


Fig. 2.4

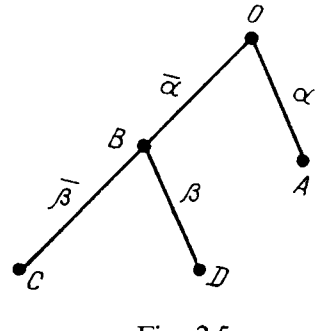


Fig. 2.5

A is the subset of trees of which branch α must be a part, B is the subset of trees in which branch α does not take part, C is the subset of trees in which branches α and β do not take part, and D is the subset of trees in which branch β must be a part, but branch α must not.

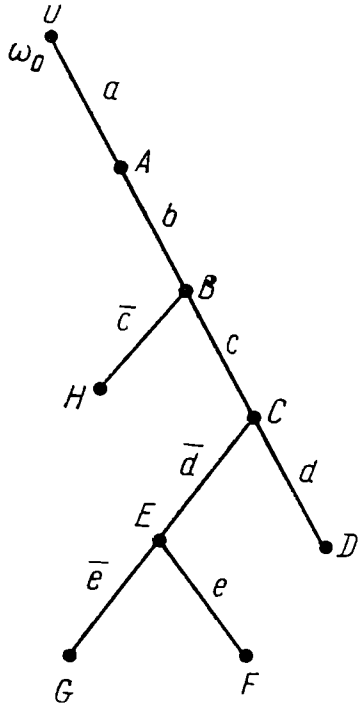


Fig. 2.6

The lower bound, ω, can be evaluated as

$$\omega = \omega_1 + \omega_2$$

where

$$\omega_1 = \min \sum a_{ij}, \qquad \omega_2 = \min \sum b_{ij} P_{ij}$$

For each hanging node, ω_1 is evaluated by the minimum length network principle. This can be done in one of two ways:

(1) from the complete graph corresponding to the subset which includes the 'mandatory' branches and does not contain forbidden branches, eliminate consecutively non-mandatory branches having the highest value of a_{ij} until a tree is generated. In doing so, do not remove the branch that makes the graph a connected one, since otherwise the constraint equations would be violated;

(2) generate a tree, starting with the mandatory branches and following with the branches with the lowest value of a_{ij}. Then to these branches attach the next branch with the minimum value of a_{ij}, originating from a bound on existing branches, then another branch with the minimum value of a_{ij}, and so on, until a tree is formed. In doing so, do not attach a branch that would turn the tree into a mesh.

For each hanging node, ω_2 is evaluated by applying the transportation technique with trans-shipments to the complete graph corresponding to a given subset. The flows associated with the mandatory branches enter the basic solution. In this way, it is possible to evaluate ω for each hanging node. In principle, branching should be continued at a node where ω is a minimum. In a modification of the network problem method* branching starts at the initial node and proceeds only to the right along the branches of the minimum length tree (Fig. 2.6). Branching proceeds in increasing order of a_{ij}, and terminates at point D corresponding to the minimum-length tree. The value of ω_0 at point O corresponds to $\min \Sigma a_{ij} + \min \Sigma b_{ij} P_{ij}$ for the complete graph. Moving downwards to the right does not affect the value of ω_1, while that of ω_2 will not change only if the corresponding new

* Lazebnik A. I. *Application of the Branch and Bound Method to the Selection of an Optimal Electric Network* (Primenenie metoda vetvey i granits dlya vybora optimalnoy elektricheskoy seti). Proceedings of USSR Academy of Sciences. "Energetika i transport", 1969, No. 2.

branch is part of tree D_2. Otherwise, ω_2 would increase, and a new transportation problem would have to be solved, which is a complicated matter. This is why ω_2 is kept constant, and ω therefore remains constant in moving downwards to the right.

Let the set at point D be a specific tree. In order to evaluate its OF, it will suffice to evaluate $\Sigma b_{ij}P_{ij}$ from the constraint equations. The value of Σa_{ij} will be the same as it is at points A and C (and equal to min Σa_{ij} for node O). As a result, the value of the objective function at point D is obtained and stored. It is to be noted that $OF_D < \omega_0$ if trees D_1 and D_2 are not coincident at point O. From point C branching proceeds to the left (with branch d forbidden). Next ω_E is evaluated at point E. In this case, the value of ω_1 is greater than it is at point C, because no minimum-length tree can exist at point E any longer. The value of ω_1 is calculated as already explained. Any branch (except branch d) is attached to branches a, b, c, provided it has the minimum value of a_{ij} and does not form a loop.

It will be necessary to recalculate ω_2 only if branch d was part of tree D_2. The value of ω_E at point E is then compared with the objective function OF_D at point D. If $\omega_E \geqslant OF_D$, no branching is done from point E, and the subset of trees at point E is not considered. Otherwise, branching proceeds downwards and to the right from point E. Branch e, having the minimum value of a_{ij}, is introduced. At point F, a tree forms, for which $\Sigma b_{ij}P_{ij}$ is evaluated from the constraint equations, and Σa_{ij} is taken to be equal to ω_1 at point E. The objective function at point E is compared with that at point D, and the lower value is stored as an optimal value of the objective function, OF_o. Now the procedure is resumed from point E; branching continues from point E to the left, towards point G, with branch e forbidden. The value of ω_G is found and compared with OF_o. If $\omega_G \geqslant OF_o$, no branching from point G is done; instead the further branching proceeds to the left from the nearest node B. Branch BH is introduced (and branch c is forbidden). Similarly, ω_H at point H is evaluated and compared with OF_o. If $\omega_H \geqslant OF_o$, no branching is continued from point H, instead, it is continued to the left from the nearest node A, and so on. If, however, branching proceeds from point H to the right, branch d may be added to the tree (the only branch forbidden is c). Whenever branching reaches the final node of the tree, the objective function is evaluated and compared with OF_o. The minimization procedure is complete when branching from point O to the left reaches a node where $\omega \geqslant OF_o$. The tree corresponding to OF_o is the global solution of the network problem.

Thus, the algorithm of a network problem solved by the bound and branch method may be stated as follows:

(1) Evaluate

$$\omega_1^{(0)} = \min \sum_{ij} a_{ij}, \qquad \omega_2^{(0)} = \min \sum_{ij} b_{ij}P_{ij}$$

$$\omega^{(0)} = \omega_1^{(0)} + \omega_2^{(0)}$$

at the initial node (point O) of the alternatives tree;

(2) proceed downwards to the right from point O along the branches of the minimum-length tree until the final node of that tree is reached. Arrange the branches in ascending order of a_{ij};

(3) at the final node, find the precise value of the objective function and store it as an optimal one, OF_0;

(4) proceed from the final node upwards and to the left to the nearest node, and from the latter to the left (a branch is forbidden). Evaluate ω for the resultant hanging node and compare the result with the optimal objective function, OF_0;

(5) if $\omega \leqslant OF_0$, proceed to the right until a tree is formed, making it certain to add branches securing the minimum-length tree, with some branches forbidden. Evaluate the objective function for the tree obtained, and compare it with OF_0. If $OF < OF_0$, store OF as a new optimal value;

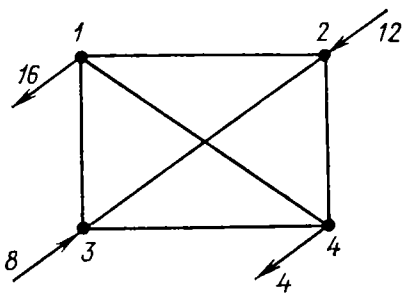


Fig. 2.7

(6) if $\omega \geqslant OF_0$ for any hanging node obtained after branching to the left, no branching should be continued from that node; instead, proceed upwards to the nearest node from which no branching to the left has been attempted;

(7) if branching from point O to the left leads to a node where $\omega \geqslant OF_0$ and if there are no more nodes lying higher and having any branching to the left, the minimization procedure is complete, and the tree answering the last value of OF_0 may be taken as being optimal.

To simplify the task, rather than to evaluate ω_2, it may be assumed everywhere that $\omega_2 = \omega_2^{(0)}$; that is, a simplified lower bound may be employed.

As will be recalled, the number of trees is n^{n-2}, where n is the number of nodes. For example, with ten nodes, there will be 10^8 trees; with twenty nodes, there will be 20^{18} trees. Thus, by excluding a node for which $\omega \geqslant OF_0$, a very great number of trees can be discarded.

By introducing the condition that if for some node $\omega < OF_0$, and

$$OF_0 - \omega < \varepsilon$$

where ε is the limit of error, then the trees associated with that node can be identified by inspection, much optimization time can be saved.

Example 2.13. Find an optimal tree in a network made up of four nodes and six branches (Fig. 2.7). The values of a_{ij} and b_{ij} are given in the table that follows.

Branch	1-2	1-3	1-4	2-3	2-4	3-4
a_{ij}	5	3	10	4	6	8
b_{ij}	0.5	0.3	1	0.4	0.6	0.8

The capacities at the nodes are specified as follows: $P_1=-16$, $P_2=+12$, $P_3=+8$, $P_4=-4$, where the '+' sign denotes generation, and the '−' sign stands for load.

To begin with, we shall find the maximum length tree, D_1, for node O of the alternatives tree.

First method. Delete branches *1-4* and *3-4* in descending order of a_{ij}. Branch *2-4* cannot be deleted because node *4* would have no branches. Therefore, we delete branch *1-2*. The resultant tree has branches *1-3*, *2-3*, *2-4* (Fig. 2.8).

Second method. Introduce branches *1-3*, *2-3*, in ascending order of a_{ij}. Branch *1-2* cannot be added, because loop *1-2-3* would be formed. Therefore, we add branch *2-4*. We get the same tree (Fig. 2.9). We find that $\omega_1^{(0)}=3+4+6=13$. To evaluate $\omega_2^{(0)}$, we should solve a transportation problem with trans-shipments.

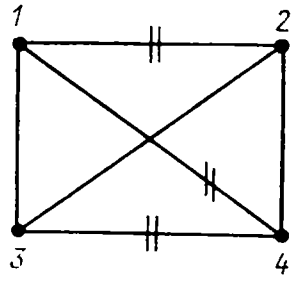


Fig. 2.8

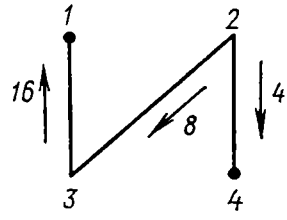


Fig. 2.9

We shall use the minimum-length tree (see Fig. 2.9) as the initial basic solution, to obtain the flow distribution as shown in the figure. The data can be cast into the following transportation tableau (array):

	1	2	3	4	a_i	
1	0	0.5	0.3	1		$-\pi_1$
2	0.5	0	8 0.4	4 0.6	12	$-\pi_2$
3	16 0.3	0.4	−8 0	0.8	8	$-\pi_3$
4	1	0.6	0.8	0		$-\pi_4$
b_j	16			4		
	π_1	π_2	π_3	π_4		

Evaluate the potentials subject to the conditions:

$$\pi_3-\pi_2=0.4$$
$$\pi_4-\pi_2=0.6$$
$$\pi_1-\pi_3=0.3$$

whence, setting $\pi_4=0$, we get

$$\pi_2=-0.6, \qquad \pi_3=-0.2, \qquad \pi_1=0.1$$

Evaluate the c_{ij}'s for the non-basic variables:

$$\bar{c}_{12}=0.5+0.1+0.6=1.2>0, \qquad \bar{c}_{13}=0.3+0.1+0.2=0.6>0$$
$$\bar{c}_{14}=1+0.1-0=1.1>0, \qquad \bar{c}_{21}=0.5-0.6-0.1=-0.2<0$$
$$\bar{c}_{34}=0.8-0.2-0=0.6>0, \qquad \bar{c}_{41}=1-0.1=0.9>0$$
$$\bar{c}_{42}=0.6+0+0.6=1.2>0, \qquad \bar{c}_{43}=0.8+0+0.2=1>0$$

Introduce the flow of branch *2-1* into the basis; $P_{21}=\theta$. Obviously, $\theta=8$. Eliminate P_{23} from the basis. As a result, we get a tree made up of branches carrying the flows $P_{21}=8$, $P_{31}=8$, and $P_{24}=4$ (Fig. 2.10).

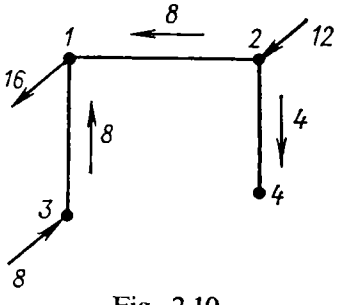


Fig. 2.10

Find the new values of the potentials subject to the conditions

$$\pi_1-\pi_3=0.3$$
$$\pi_1-\pi_2=0.3$$
$$\pi_4-\pi_2=0.6$$

Setting $\pi_4=0$, we get

$$\pi_2=-0.6, \qquad \pi_1=-0.1, \qquad \pi_3=0.4$$

Evaluate the $\bar{c}_{ij}$'s for the non-basic variables:

$$\bar{c}_{12}=0.5-0.1+0.6=1>0, \qquad \bar{c}_{13}=0.3-0.1+0.4=0.6>0$$
$$\bar{c}_{14}=1-0.1-0=0.9>0, \qquad \bar{c}_{23}=0.4-0.6+0.9=0.2>0$$
$$\bar{c}_{32}=0.4-0.4+0.6=0.6>0, \qquad \bar{c}_{34}=0.8-0.4-0=0.4>0$$
$$\bar{c}_{41}=1+0+0.1=1.1>0, \qquad \bar{c}_{42}=0.6+0+0.6=1.2>0$$
$$\bar{c}_{43}=0.8+0+0.4=1.2>0$$

that is, all $\bar{c}_{ij}>0$. The optimal tree appears in Fig. 2.10.

Evaluate $\omega_2^{(0)}$ (see Fig. 2.10):

$$\omega_2^{(0)}=8\times0.3+8\times0.5+4\times0.6=8.8$$

For the initial node,

$$\omega^{(0)}=13+8.8=21.8$$

Branch to the right in ascending order of a_{ij}. We obtain points *A*, *B* and *C* (Fig. 2.11).

To evaluate OF_C at point C, we should find

$$\sum a_{ij}=\omega_1^{(0)}=\min \sum a_{ij}=13$$

and from the flow distribution (see Fig. 2.9)

$$\sum b_{ij}P_{ij}=16\times 0.3+8\times 0.4+4\times 0.6=10.4$$

Then,

$$OF_C=13+10.4=23.4$$

This value is stored as an optimal one.

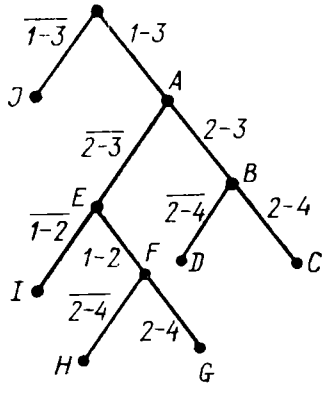


Fig. 2.11

Branching is then carried on from node B to the left as far as point D (with branch *2-4* forbidden). Find the lower bound for node D, that is, ω_D. The value of ω_{1D} is found by attaching branch *3-4* to mandatory branches *1-3* and *2-3*. Branch *2-4* is forbidden and branch *1-2* would form a loop. Therefore,

$$\omega_{1D}=3+4+8=15$$

Tree D_2 is found by solving a transportation problem. Since point D is associated with only two trees *(1-3, 2-3, 3-4* and *1-3, 2-3, 1-4)*, it is simpler to evaluate $\Sigma b_{ij}P_{ij}$ for these trees and select that which has the minimum value of $\Sigma b_{ij}P_{ij}$.

For the first tree

$$\sum b_{ij}P_{ij}=16\times 0.3+12\times 0.4+4\times 0.8=12.8$$

and for the second,

$$\sum b_{ij}P_{ij}=20\times 0.3+12\times 0.4+4\times 1=14.8>12.8$$

Therefore,

$$\omega_D=15+12.8=27.8>OF_o=23.4$$

When use is made of a simpler evaluation of the lower bound, where only ω_1 is determined and $\omega_2=\omega_2^{(0)}$, we get

$$\omega_D\approx 15+8.8=23.8>23.4$$

Thus, no further branching from node D is advisable, and the node is discarded along with the two associated trees.

Then branching is continued to the left from node A lying above towards point E (with branch *2-3* forbidden). Point E has five trees associated with it. By a similar reasoning, it can be proven that for this point tree D_1 is coincident with tree D_2 *(1-3, 1-2, 2-4)*.

It is to be noted that branch *2-4* formerly forbidden has not been added to tree D_1. From the computations

$$\omega_{E1}=3+5+6=14$$
$$\omega_{E2}=8\times 0.5+8\times 0.3+4\times 0.6=8.8$$
$$\omega_{E3}=14+8.8=22.8<23.4$$

we conclude that we should continue branching from point E.

From point E, branching should proceed to the right towards point F (branch *1-2* is added). With branch *2-3* forbidden, branch *1-2* has the least value of a_{ij}. Point F has three trees associated with it, D_1 being coincident with D_2 (branches *1-3*, *1-2*, *2-4*); as before, $\omega_{F1}=14$, $\omega_{F2}=8.8$ and $\omega_F=22.8<23.4$.

From point F, branching now proceeds to the right, towards point G (by adding branch *2-4* formerly forbidden). For tree G

$$OF_G=14+8.8=22.8<23.4$$

that is, tree G is now optimal, and the new value of $OF_0=22.8$ is stored. Branching is carried on from point G to the left in the direction of point H (with branch *2-4* forbidden). Point H has two trees associated with it. At point H, trees D_1 and D_2 coincide and correspond to a tree with branches *1-3*, *1-2* and *1-4*. Therefore,

$$\omega_{H1}=3+5+8=16$$
$$\omega_{H2}=12\times0.5+4\times0.3+4\times0.8=10.4$$
$$\omega_H=16+10.4=26.4>22.8$$

Using an approximate evaluation,

$$\omega_H\approx16+8.8=24.4>22.8$$

Thus, subset H (two trees) may be discarded.

Branching from point E to the left (with branch *1-2* forbidden) leads to point I which likewise has two trees associated with it. Analysis shows that tree D_1 has branches *1-3*, *3-4* and *2-3* (branches *1-2* and *2-3* are forbidden). Therefore,

$$\omega_{I1}=3+8+6=17$$

Tree D_2 has branches *1-3*, *1-4* and *2-4*, and so

$$\omega_{I2}=12\times0.6+8\times0.3+8\times1=17.6$$
$$\omega_I=17+17.6=34.6>22.8$$

With an approximate evaluation,

$$\omega_I\approx17+8.8=25.8>22.8$$

Subset I is thus discarded, and branching proceeds to the left from point O to point J.

With branch *1-3* forbidden, set J contains eight trees radiating as shown in Fig 2.12. Tree D_1 has branches *1-2*, *2-3* and *2-4*, and so

$$\omega_{J1}=5+4+6=15$$

Tree D_2 has branches *1-2*, *2-3* and *3-4*, and so

$$\omega_{J2}=16\times0.5+4\times0.4+4\times0.8=12.8$$
$$\omega_J=15+12.8=27.8>22.8$$

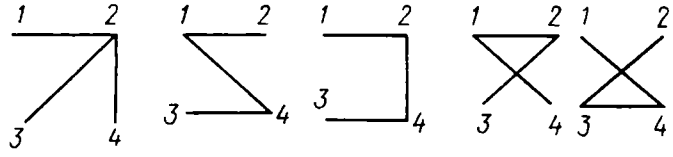


Fig. 2.12

With an approximate evaluation,

$$\omega_J = 15 + 8.8 = 23.8 > 22.8$$

This completes the computation, because no further branching is possible. The optimal tree is at point *G* (with branches *1-3*, *1-2* and *2-4*, see Fig. 2.10), that is, a tree answering min $\Sigma b_{ij} P_{ij}$.

It is to be noted that the total number of trees is $4^2 = 16$. The precise value of the objective function has been computed for only two trees *(C* and *G)*. The lower bound has been evaluated for six points *(D, E, F, H, I* and *J)*, of which four nodes have been discarded, namely *D* (two trees), *H* (two trees), *I* (two trees) and *J* (eight trees). Two nodes, *E* and *F*, have not been discarded, and branching continued therefrom.

A total of eight steps has been made, with the optimal solution found at the fifth step.

2.12. Cost-of-Loss Block

The purpose of the cost-of-loss block is to evaluate the reliability of power supply. The block determines the expected cost of loss of supply due to faults or breakdowns at the stations, with or without account of the probabilistic character of hydrolicity. In a simplified form, this block is examined elsewhere in the book.

When the probabilistic characteristics of the hydro stations are ignored, the following probability distributions are sampled: for loss of generation, with the sampling interval, ε, chosen arbitrarily *(*in *MW)*, and for power deficiency by combining samples from the loss-of-generation probability distribution and the total load on the power system. Then, the expectation of power deficiency (failure to meet demand) and the respective expectation of the cost of power not supplied (cost of supply lost) are computed.

In order to determine the expectation of the cost of supply lost, it is necessary to know the probability distribution of the operating state, p, and the abnormal state, q, of each generating unit within the system, the load duration curve, and the specific cost of supply lost (roubles per kWh).

2.13. Integer Linear Programming and Mathematical Formulation of an Expansion Optimization Problem

In some problems, one seeks the minimum of a linear function, where the constraint equalities and nonequalities are linear, but the variables appearing in the final solution must be whole numbers. Of many methods employed in tackling such problems, we shall examine only the first algorithm of Gomory's method.

The rationale of this method is that after the objective function has been minimized on the assumption that the unknowns are continuous and after an appropriate optimal basic solution has been obtained, an additional basic unknown is introduced, and also an appropriate constraint equation of a form such that one of the basic variables would remain a whole number as the additional unknown is eliminated from the basis. Removal of such an equation is called *cutting*. The solution thus found may well be non-basic. To secure that the next basic variable be a whole number, a new unknown is introduced, and so on.

As will be recalled, Gomory's method converges to the sought solution, provided all constraint equations, including that for z where the basic unknown is $-z$, are arranged in a certain order and that the basic variables are made whole numbers as a result of cutting in the same order. Besides, if additional unknowns are removed from the non-basic ones, no cut solution is introduced in the constraint equations any longer.

Example 2.14. Minimize the function

$$z=-x_1-x_2$$

subject to

$$\begin{aligned} 2x_1+11x_2 &\leqslant 38 \\ x_1+x_2 &\leqslant 7 \\ 4x_1-5x_2 &\leqslant 5 \end{aligned}$$

and also that x_1, x_2 and z must be whole numbers.

Convert the constraint inequalities to equalities. As a result, we get a system of equations:

$$\begin{aligned} 2x_1+11x_2+x_3 &= 38 \\ x_1+x_2+x_4 &= 7 \\ 4x_1-5x_2+x_5 &= 5 \\ -x_1-x_2-z &= 0 \end{aligned}$$

The basic solution may be written as:

$$x_3=38, \quad x_4=7, \quad x_5=5, \quad x_1=x_2=0, \quad z=0$$

Introduce x_1 into, and eliminate x_5 from, the basis. From the third equation we get

$$x_1=5/4+5x_2/4-x_5/4$$

Substituting x_1 in the other equations gives:

$$\begin{aligned} x_1 - \tfrac{5}{4}x_2 \quad + \tfrac{1}{4}x_5 \quad &= \tfrac{5}{4} \\ \tfrac{27}{2}x_2 + x_3 - \tfrac{1}{2}x_5 \quad &= \tfrac{71}{2} \\ \tfrac{9}{4}x_2 + x_4 - \tfrac{1}{4}x_5 \quad &= \tfrac{23}{4} \\ -\tfrac{9}{4}x_2 \quad + \tfrac{1}{4}x_5 - z &= \tfrac{5}{4} \end{aligned}$$

Introduce x_2 into, and eliminate x_4 from, the basis. From the third equation we get

$$x_2 = 23/9 - 4x_4/9 + x_5/9$$

Substituting x_2 in the other equations gives:

$$\begin{aligned} x_1 \quad + \tfrac{5}{9}x_4 + \tfrac{1}{9}x_5 \quad &= \tfrac{40}{9} \\ x_2 \quad + \tfrac{4}{9}x_4 - \tfrac{1}{9}x_5 \quad &= \tfrac{23}{9} \\ x_3 - \ 6x_4 + \ x_5 \quad &= 1 \\ x_4 \quad - z &= 7 \end{aligned}$$

The basic solution ($x_1 = 40/9$, $x_2 = 23/9$, $x_3 = 1$, $x_4 = x_5 = 0$) thus found is optimal, and $z = -7$.

To make x_1 a whole number, we introduce an additional unknown, x_6, defined by the cut equation as

$$x_6 = 5x_4/9 + x_5/9 - 4/9$$

The right-hand side of the cut equation contains the unknowns which enter the first equation with coefficients equal to the fractional excess over the nearest whole number. The respective fractional excess of the free term has the '$-$' sign. If we eliminate x_6 from the basis, that is, if we equate it to zero, then substituting the cut equation

$$0 = 5x_4/9 + x_5/9 - 4/9$$

in the first equation makes x_1 a whole number.

The new system of cut equations will be written as

$$\begin{aligned} x_1 \quad + \tfrac{5}{9}x_4 + \tfrac{1}{9}x_5 \quad &= \tfrac{40}{9} \\ x_2 + \quad \tfrac{4}{9}x_4 - \tfrac{1}{9}x_5 \quad &= \tfrac{23}{9} \\ x_3 - \ 6x_4 + \ x_5 \quad &= 1 \\ -\tfrac{5}{9}x_4 - \tfrac{1}{9}x_5 + x_6 \quad &= -\tfrac{4}{9} \\ x_4 \quad - z &= 7 \end{aligned}$$

Introduce x_6 into, and eliminate x_4 from, the basis. The unknown to be introduced into the basis is decided by the minimum value of b_i/a_{is}, that is

$$\min\left|\frac{-4/9}{-5/9}, \quad \frac{-4/9}{-1/9}\right| = \min\left|\frac{4}{5}, \quad 4\right| = \frac{4}{5}$$

After the equation is transformed, the unknown to be introduced, x_4, can be found from the last constraint equation:

$$x_4 = 9x_6/5 - x_5/5 + 4/5$$

Substituting x_4 in the remaining equations of the system yields a new system of equations:

$$\begin{aligned}
x_1 \qquad\qquad\quad + \quad x_6 \quad &= 4 \\
x_2 \quad -\frac{1}{5}x_5 + \frac{4}{5}x_6 \quad &= \frac{11}{5} \\
x_3 + \quad \frac{11}{5}x_5 - \frac{54}{5}x_6 \quad &= \frac{29}{5} \\
x_4 + \quad \frac{1}{5}x_5 - \frac{9}{5}x_6 \quad &= \frac{4}{5} \\
-\frac{1}{5}x_5 + \frac{9}{5}x_6 - z &= \frac{31}{5}
\end{aligned}$$

In this system, x_1 is a whole number.

To make the unknown next in order, x_2, also a whole number, we introduce a new cut equation

$$x_7 = -x_5/5 + 4x_6/5 - 1/5$$

and obtain a new system of equations:

$$\begin{aligned}
x_1 \qquad\qquad\quad + \quad x_6 \quad &= 4 \\
x_2 \quad -\frac{1}{5}x_5 + \frac{4}{5}x_6 \quad &= \frac{11}{5} \\
x_3 + \quad \frac{11}{5}x_5 - \frac{54}{5}x_6 \quad &= \frac{29}{5} \\
x_4 + \quad \frac{1}{5}x_5 - \frac{9}{5}x_6 \quad &= \frac{4}{5} \\
\frac{1}{5}x_5 - \frac{4}{5}x_6 + x_7 &= -\frac{1}{5} \\
-\frac{1}{5}x_5 + \frac{9}{5}x_6 - z &= \frac{31}{5}
\end{aligned}$$

Eliminate x_7 from, and introduce x_6 into, the basis.

From the last constraint equation, we find

$$x_6 = 1/4 + x_5/4 + 5x_7/4$$

Substituting x_6 in all the other equations yields a new system of equations:

$$\begin{aligned}
x_1 \quad & +\frac{1}{4}x_5+\frac{5}{4}\,x_7 = \frac{15}{4}\\
x_2 \quad & + x_7 = 2\\
x_3 \quad & -\frac{1}{2}x_5-\frac{27}{2}x_7 = \frac{17}{2}\\
x_4 \quad & -\frac{1}{4}x_5-\frac{9}{4}\,x_7 = \frac{5}{4}\\
& \frac{1}{4}x_5+\frac{9}{4}\,x_7-z=\frac{23}{4}
\end{aligned}$$

The equations for the basic variable x_6 is not included in the above system in accordance with Gomory's algorithm.

Now x_2 is a whole number, but x_1 is not. In accordance with Gomory's algorithm, we are to carry out cutting so as to make x_1 a whole number again.

We introduce a cut equation

$$x_8=x_5/4+x_7/4-3/4$$

Now eliminate x_8 from, and introduce x_5 into, the basis. Evaluate x_5 from the cut equation

$$x_5=3-x_7+4x_8$$

Substituting x_5 in the other equations gives a new system:

$$\begin{aligned}
x_1 \quad & + x_7+ x_8 = 3\\
x_2 \quad & + x_7 = 2\\
x_3 \quad & -13x_7-2x_8 = 10\\
x_4 \quad & - 2x_7- x_8 = 2\\
x_5 \quad & + x_7-4x_8 = 3\\
& x_7+ x_8-z= 5
\end{aligned}$$

The solution thus obtained is optimal and contains only whole numbers:

$$x_1=3, \quad x_2=2, \quad x_3=10, \quad x_4=2, \quad x_5=-3, \quad z=-5$$

This solution contains six non-zero terms; that is, this solution is not a basic one.

Thus, it has taken two steps to obtain an optimal basic solution with non-integer variables and three more steps to obtain an optimal integer solution.

In the general case, a power system expansion optimization problem may be stated as an integer programming problem.

Let there be M existing stations, that is, $m=1, 2, \ldots, M$, and N sites for new stations, that is, $n=1, 2, \ldots, N$, and there are l_n alternative designs for each new station, each specifying different station parameters (generating-unit capacity, steam conditions, etc.), that is, $l_n=1, 2, \ldots, L_n$. Also, there are S alternative designs for power transmission lines, that is, $j=1, 2, \ldots, S$. We further let the discounted cost of construction per year for a new nth station under an lth alternative design be as follows:

$$\begin{aligned} C&=0 && \text{for} \quad t<t_c \\ C&=a_{nl}+b_{nl}\Delta P_{nlt_c} && \text{for} \quad t=t_c \\ C&=b_{nl}\Delta P_{nlt} && \text{for} \quad t>t_c \end{aligned}$$

where t_c is the commissioning year and ΔP_{nlt} is the capacity commissioned in the year t at the nth station built under the lth alternative design.

We introduce Boolean variables Y_{nlt} which take on whole values, 0 or 1. Then the discounted cost of construction for every year of the study period for all new stations can be written as

$$\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{L_n} Y_{nlt}a_{nlt}+\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{L_n} b_{nlt}\Delta P_{nlt}$$

subject to

$$\sum_{t=1}^{T}\sum_{l=1}^{L_n} Y_{nlt}\leqslant 1, \qquad \sum_{t=1}^{T}\sum_{l=1}^{L_n} Y_{nlt}P_{nl\max}\geqslant\sum_{t=1}^{t}\Delta P_{nlt}$$

In view of the first constraint, Y_{nlt} can take on a unity value only for one specific value of t, $t=t_c$, and for only one specific value of l_n. From the second constraint it follows that for all $t<t_c$ the left-hand side of the inequality is zero. That is, at $t<t_c$, no generating capacity can be commissioned at the nth station in any year. If $Y_{nlt}=1$ (at $t=t_{cn}$ and a specific value of l_n), then for the year t_{cn} the cost will be $a_{nl}+b_{nl}\Delta P_{nlt_c}$. Some capacity must be commissioned at the nth station in that year by all means, or else $Y_{nlt}=0$, and there would be no sense in building that station at all. Also, since $Y_{nlt}=0$ for all other l's, all $\Delta P_{nlt}=0$ for all other values of l in view of the second constraint. That is, the generating capacity at the nth station will be commissioned only under the optimal design l and only starting with the optimal commissioning year t_c.

The fuel cost for all stations may be estimated approximately as follows:

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{h_{m\min}}^{h_{m\max}} c_{mt}^{h}P_{m}^{h}+\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l=1}^{L_n}\sum_{h_{n\min}}^{h_{n\max}} c_{nlt}^{h}P_{nlt}^{h} \tag{2.75}$$

where c_{mt}^{h} and c_{nlt}^{h} are the specific operating costs (including fuel) for a definite number of operating hours (h_m and h_n) such that the number of hours can be changed from $h_{\min}$ to $h_{\max}$ only in steps. For each alternative design there will be a particular value of c_{nlt}^{h} differing from those of the other alternatives. The constraints are stated so as to supply the peak load and deliver the required quantity of energy.

For the power transmission lines, the discounted costs of construction and operation can likewise be specified for different alternative designs as follows:

$$\left.\begin{aligned} C_{trans} &= 0 && \text{for} \quad t < t_c \\ C_{trans} &= a_{ijl} + b_{ijl} P_{ijlt} && \text{for} \quad t = t_c \\ C_{trans} &= C_{ijl} P_{ijl} && \text{for} \quad t > t_c \end{aligned}\right\} \tag{2.76}$$

where C_{ijl} = operating costs including energy lost (for existing lines this equation is highly approximate and applies to cases where the power flows in the lines are constant)

P_{ijl} = maximum power flow equal to the maximum handling capacity

Apart from the constraints

$$\sum_{t=1}^{T} Y_{ijlt} \leqslant 1$$

and

$$\sum_{t=1}^{T} \sum_{l=1}^{L} Y_{ijlt} P_{ij\,\max} \geqslant P_{ijlt}$$

there are constraints on the balance of power at the nodes (arising from Kirchhoff's current law). The unknowns in this problem are Y_{nlt}, ΔP_{nlt}, h_{mt}, h_{nt}, Y_{ijlt}, and P_{ijl}. The terms Y_{nlt} and Y_{ijlt} are integer quantities and can be a 0 or a 1. Because of the great number of unknowns and constraint equations involved, a rigorous solution of this problem cannot yet be obtained by integer solution techniques, as it would overtax any computer.

REFERENCES TO CHAPTER 2

1. Dantzig G. B. *Linear Programming and Extensions.* Princeton (N. G.), Princeton Univ. Press, 1963.
2. Markovich I. M. and Sharnopolsky B. P. *On a Likely Application of the Generalized Transportation Problem to Power-System Optimization* (Ob odnoy vozmozhnosti ispolzovaniya obobshchennoy transportnoy zadachi dlya opredeleniya optimalnoy elektricheskoy seti). Energetika i transport, 1967, No. 5.
3. Korbut A. A. and Finkelshtein Yu. Yu. *Discrete Programming* (Discretnoe programmirovanie). Nauka, 1969.
4. Kokhov V. I. *Adjustment for Constant Cost Items in the Optimization of Power-System Expansion Projects* (Uchet postoyannykh sostavlyayushchikh pri optimizatsii elektroenergeticheskikh sistem). Energetika i transport, 1968, No. 3.
5. Markovich I. M. *Optimization of Power-System Operation* (Optimizatsii rezhimov energeticheskikh sistem). Moscow, Power Institute Press, 1967.
6. Markovich I. M. and Rokotyan S. S. *Optimization and Estimation Models for the Selection of an Optimal Expansion Scheme for Electric Power Systems* (Optimizatsionnye i otsenochnye modeli dlya vybora optimalnogo razvitiya elektroenergeticheskikh sistem). Power System Press, Siberian Branch of the USSR Academy of Sciences, 1970.
7. Motskus I. B. *Multi-Extremal Problems in Design* (Mnogoextremalnye zadachi v proektirovanii). Nauka, 1962.

CHAPTER 3

CRITERION ANALYSIS OF THE TECHNICAL AND ECONOMIC PROBLEMS OF THE POWER INDUSTRY

3.1. Selection Criterion for an Optimal Alternative*

Most engineering projects, irrespective of their field (be it in civil engineering, in the power industry, or elsewhere) and of their scale, can be carried out in several ways, each using an approach of its own, a different technology, and different amounts of materials and manpower. For example, in examining alternative plans for a power system, we may choose either a thermal or a hydro station as a source of electric energy. As a rule, any alternative design differs from any other in a great many technical features. Yet, there may be cases where the alternatives under consideration will only differ in one, most decisive factor, which may be steam pressure or temperature, line cross-sectional area, transformer substation capacity, motor speed, etc.

In giving his preference to a particular alternative, the engineer does so because he believes that the selected plan is better than the rest in some way. The indicators used in selecting from among several alternative designs are called *optimality criteria.* They exist always, provided the alternative plans have been drawn up on a sound scientific footing. Otherwise, no founded selection would be possible. The alternative complying with the adopted criteria is taken to be an optimal one as regards the criteria used.

It would appear that an optimal alternative should be that which embodies the latest advances in science and technology. However, a state-of-the-art design will not necessarily be an optimal one at a particular time. It is more correct to give preference to an alternative which yields the desired technical payoff at a minimum cost or the maximum technical payoff at a specified fixed cost. This is the reason why it is customary to compare alternative plans for their technical and economic advantages and limitations.

The selection of an economically sound alternative can only be done on the basis of an aggregated economic indicator. This should preferably be the discounted sum of capital and operating costs, defined as

$$C = p_{st} I + O \tag{3.1}$$

---

* See also *Electrical Systems*, vol. II, ed. by V. A. Venikov. *Electrical Networks*, Chap. 7. Higher School Publishers, 1971. In the further references to this publication, only the volume and chapter will be quoted.

where p_{st} = standard discount rate
I = investments
O = yearly operating costs

Currently, it is assumed that an optimal alternative design is that which has the lowest discounted sum of capital and operating costs, with all other conditions being equal.

In some cases, it may be convenient to use as optimality criteria some physical quantities such as weight, volume, loss of energy, etc. For example, the minimum energy lost may serve as an optimality criterion in the selection of an economically sound load factor for power transformers.

3.2. Technical-Economic Models

In a technical and economic study, it is important to take a proper account of its two aspects, qualitative and quantitative. Apart from being interrelated, they serve the same purpose: provide deep insight into the study subject. On the other hand, each aspect has features of its own.

Qualitative analysis is the first step in forming an idea about the study object. It is very important when the investigator comes to stating what he thinks about the physical and economic features of the study object, and it is closely related to abstraction, idealization and modelling. For all their importance, however, theoretical ideas become concrete only when they are coached in quantitative terms. It may be said that the logical nets of qualitative analysis are coarse but strong, whereas, the mathematical nets of quantitative analysis are woven fine but they are too weak to hold the problem under study unless it has been wrapped up in the logical nets of qualitative analysis.

In getting down to an analysis of a problem, the investigator should above all define the qualitative boundary of the study object, that is, isolate it from the surroundings. The point is that real objects of study possess a multitude of properties some of which have to be sacrificed or ignored during an investigation. For example, in determining voltage loss in L.V. overhead lines it is usual to neglect capacitive admittance and conductance, the rise in resistance with rise in conductor temperature caused by the flow of current and, the more so, by solar radiation. This listing of immaterial properties might be continued *ad infinitum*. The investigator selects a finite number of properties and discards what he thinks to be irrelevant to the problem at hand, thereby idealizing the study object. The collection of properties thus chosen defines the qualitative boundary of the study.

Any object of technical and economic analysis is characterized by a multitude of parameters which may be divided into the following categories: technical, economic and technico-economic. Examples of technical parameters are voltage, power, speed and weight. They describe the physical properties of the study object. The economic parameters include discounted costs, capital costs and operating or running costs: they describe the study

object money-wise. Technico-economic parameters stand mid-way between the first two categories. They are exemplified by the cost of the kilowatthour of energy generated, specific costs of station construction.

According to the role the design parameters ($a_1, a_2, \ldots, a_\gamma$) play in technical and economic analysis, they may be divided into two classes.

The first class of design parameters ($a_1, a_2, \ldots, a_s$) embraces those which do not vary numerically in the course of problem solving. For convenience in analysis and in order to reduce notation, these parameters, where possible, should preferably be combined in a generalized type of parameters such as

$$A_i = \varphi_i(a_1, a_2, \ldots, a_s) \quad i = \begin{vmatrix} m \\ 0 \end{vmatrix} \tag{3.2}$$

and referred to as *constants.*

The second class of design parameters includes those which vary numerically in the course of problem-solving, and quite aptly they are referred to as *variable parameters* or, simply, *variables.* They are designated as follows:

$$x_1 = a_{s+1}, \quad x_2 = a_{s+2}, \quad \ldots, \quad x_n = a_\gamma \tag{3.3}$$

Of course, this division of parameters into constants and variables is arbitrary, because one and the same parameter may be a constant known prior to problem-solving or an unknown variable in a different investigation. For example, in determining losses in a power transmission line, the conductor cross-sectional area F may be taken as a specified constant. If, on the other hand, the investigator is concerned with the effect of conductor cross section on costs, F will be a variable.

This division of parameters is arbitrary also because the constants A_i are not sometimes true constants, but take a range of values due to inaccuracies in the original data. This is especially true of the constants which are dependent on economic or technico-economic parameters.

For a quantitative analysis, it is not sufficient to know design parameters alone. Already at the stage of qualitative analysis, it is important to establish the functional relationship between the parameters. It is usually sought to present this relationship as a system of mathematical equations which are then treated as a technico-economic model of the study object.

In comparison with purely technical or purely economic models, a distinction of such models is that one of the basic equations is that describing the discounted sum of costs

$$C = F(A_0, A_1, \ldots, A_m, x_1, x_2, \ldots, x_n) \tag{3.4}$$

which relates the costs C to the constants A_i and the variables x_j.

The form that Eq. (3.4) may actually take depends on the technical and economic features of the study object and the goals that the engineer strives to achieve by his analysis. For example, in order to determine the econo-

mical cross-sectional area of conductors for a power transmission line the equation of discounted costs is usually written as

$$C=(p_{st}+a/100)K_0'l+(p_{st}+a/100)K_0''lF+3I_{\max}^2b\tau\varrho l/F \tag{3.5}$$

where p_{st} = standard discount rate
a = total per cent depreciation and repair deductions
K_0' = construction cost per kilometre of line, independent of conductor cross section
K_0'' = construction cost per kilometre of line, dependent on conductor cross section
l = length of the line
F = cross-sectional area of the conductor
$I_{\max}$ = maximum current in the line
b = cost of energy lost per kWh
τ = time of loss
ϱ = conductor resistivity

It is seen from Eq. (3.5) that the cost of the line is the sum of terms each of which takes care of a particular effect. For example, the term $3I_{\max}^2b\tau\varrho l/F$ represents the cost of energy lost. Each of the design parameters included in the model may be looked upon as an independent variable which takes on a new numerical value each time the input data are changed. From the view-point of the problem being solved, that is, the economic analysis of the conductor cross section to be used, what matters is not so much changes in the individual parameters, as those in a certain set of parameters. For example, if the cost per kWh of the energy lost b be reduced in the same proportion as the time of loss τ is increased, the overall cost of energy lost and, indeed, the discounted sum of costs will remain unaffected.

Equation (3.5) has been derived in order to determine the economically justified conductor cross-sectional area, F, which is the objective parameter (one being optimized) x, in this problem. All the remaining parameters may be lumped together into the generalized constants A_i:

$$\left.\begin{aligned} A_0&=(p_{st}+a/100)K_0'l\\ A_1&=(p_{st}+a/100)K_0''l\\ A_2&=3I_{\max}^2b\tau\varrho l \end{aligned}\right\} \tag{3.6}$$

Then Eq. (3.5) may be generalized as follows:

$$C=A_0+a_1F+A_2F^{-1} \tag{3.7}$$

Cost equations can similarly be derived for other study objects. Generalizing Eq. (3.7) to the case where the number of objective variables (those being optimized) is n, and the number of effects taken into account is m, the equation of discounted costs may be written as

$$C=C_0+\sum_{i=1}^{m}A_i\prod_{j=1}^{n}x_j^{\alpha_{ij}} \tag{3.8}$$

From analysis of the above generalized discounted cost equation we can above all conclude that it is the sum of component costs

$$C=C_0+\sum_{i=1}^{m}C_i \tag{3.9}$$

The equation reflects the fact that the costs of a system as a whole are the sum of the costs of the system elements. In turn, each term in the cost equation

$$C_i=A_i\prod_{j=1}^{n}x_j^{\alpha_{ij}} \tag{3.10}$$

is the product of the constant A_i and the variables x_j raised to power α_{ij}. This form of presentation for the cost of each system element is widely used in practical technical and economic analysis of electric power installations. Sometimes, the cost equation takes a more elaborate form:

$$C_i=\psi_i(A_i, x_j), \quad i=\overline{1,m}, \quad j=\overline{1,n} \tag{3.11}$$

These are, however, special cases met with seldom in practice, and they will not be discussed any more.

The term C_0, which is independent of changes in the variables x_j, is usually called the *fixed cost item.* Since it has no effect on the economically reasonable values of the variables x_j, it is often dropped to simplify Eq. (3.8).

However, C_0 has a marked effect on the total cost C. For example, in the case of a power transmission line it may account for as much as 70 to 80% of the total cost [see Eq. (3.7)]. Therefore, in determining the absolute value of the costs C, it is mandatory to include C_0 in Eq. (3.8).

The numerical value of the exponents α_{ij} depends on the physical laws governing the study object and the methods adopted for the approximation of input data. As will be shown shortly, α_{ij} plays an important role in quantitative analysis based on the theory of similarity. Assume that these exponents may take any real values in the range

$$-\infty<\alpha_{ij}<\infty \tag{3.12}$$

To build a complete technical and economic model of the study object, we need a series of additional equations which come about from analysis of the technical constraints imposed on changes in the variables x_j.

From purely mathematical considerations, the variables x_j may formally take any values, positive, zero, negative and complex. However, in the light of the physical character of the process under study, we conclude that these variables can only fall in the positive range, in which case

$$x_j>0 \tag{3.13}$$

For practical problems of the power industry, the range on which x_j is definable is narrowed still more. For example, the cross-sectional area of

aluminium conductors for L.V. power transmission lines can range from a minimum value, $F_{\min}$, decided by mechanical strength, to a maximum value, $F_{\max}$, decided by ease of installation and design of towers. This is why Eq. (3.13) is replaced by the condition

$$x_{j\min} \leqslant x_j \leqslant x_{j\max} \tag{3.14}$$

However, this constraint needs qualification, too. The point is that most of the parameters describing electric power installations may take only discrete values limited either by the range of products manufactured or by appropriate standards. For example, the range of cross sections for aluminium-steel conductors includes the following sizes: 10, 16, 25, 35, 50, 70, 95, 120, 150, 185, 240, 300, 400, 500, 600, and 700 mm^2. This is why for many variables the valid range of values as described by Eq. (3.14) should be treated as one comprising only discrete values.

The constraints imposed on changes in the variables often take the form

$$\sum_{s=1}^{r} B_{ts} \prod_{j=1}^{n} x_j^{a} tsj \leqslant 1, \quad t = \overline{|}\begin{matrix} k \\ 1 \end{matrix} \tag{3.15}$$

Subsequently referred to as functional, this form of constraints stems from a number of technical requirements. For example, the cross-sectional area of conductors for L.V. transmission line should be chosen so that the voltage drop and the temperature of the conductor material should not exceed certain limits.

With all imposed constraints taken into account, the technical and economic model of the installation under study is usually written as

$$\left.\begin{aligned} C &= C_0 + \sum_{j=1}^{m} A_i \prod_{j=1}^{n} x_j^{a} ij \\ \sum_{s=1}^{r} B_{ts} \prod_{j=1}^{n} x_j^{a} tsj &\leqslant 1, \quad t = \overline{|}\begin{matrix} k \\ 1 \end{matrix} \\ x_{j\min} &\leqslant x_{j\max}, \quad j = \overline{|}\begin{matrix} n \\ 1 \end{matrix} \end{aligned}\right\} \tag{3.16}$$

3.3. Basic Tasks of Technical-Economic Analysis

We set out to formulate the tasks of technical and economic analysis. For ease of visualization, we shall use the simplest cost equation of the form of Eq. (3.7) which may be written as

$$C = C_0 + A_1 x + A_2 x^{-1} \tag{3.17}$$

With this equation, the material can readily be illustrated with geometrical construction on a plane (Fig. 3.1). The values of x will be laid off as abscissa and the values of discounted cost C as ordinate. If the constants C_0, A_1 and A_2 and the variable x_1' are specified in advance, the value of C_1' can be found by Eq. (3.17). From the view-point of technical-economic ana-

lysis, a specific alternative design of the installation under study can well be defined by a particular set of constants, variable and costs. On the xOC plane, this alternative is represented by point A.

If, with the same values of the constants, we cause the variable x to take a new value, x_1', there will be a new alternative design, B, for which the cost will be C_1'. By varying x from x_α to x_β, we shall obtain a multitude of alternatives, $M_1(x)$, represented by curve I in the plot of Fig. 3.1.

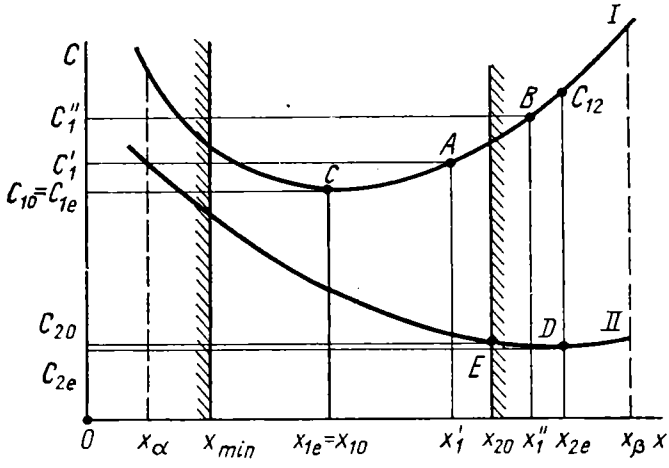


Fig. 3.1

Similarly, we can obtain a multitude of alternatives with another set of constants C_0, A_1 and A_2. This set of alternatives $M_2(x)$, is represented by curve II.

Thus, changes in the set of constants generate a new set of alternatives. In contrast, assigning a particular value to the variable x uniquely defines a particular alternative in a given set of alternatives. This example shows how different the constants and the variable are in their effects.

One of the tasks that technical-economic analysis seeks to achieve, given a particular set of initial data C_0 and A_i, is to single out of the set of alternatives $M(x)$ such a value of x_e for which the discounted cost would be a minimum, C_e. These values, x_e and C_e, together form the so-called *economical vector of variables*, associated with an economical alternative. In the plot of Fig. 3.1, economical alternatives for different sets of initial data are represented by points C and D.

Another task of technical-economic analysis is to determine the relative contributions from the constituent terms of the discounted cost equation, which may be called technico-economic proportioning. This problem is often encountered when the available resources are to be allocated between the individual installations of a power system. This problem can readily be solved once an economical alternative is known. Thus, for Eq. (3.17), the relative cost, π_{ie}, per term, can be found from the relations:

$$\pi_{0e}=C_0/C_e, \qquad \pi_{1e}=A_1x_e/C_e, \qquad \pi_{2e}=A_2x_e^{-1}/C_e \tag{3.18}$$

It often happens that inaccurate or assumed input or initial data have to be relied upon in technical and economic analysis. This is especially so when a new type of plant or equipment is being designed. It is therefore important to investigate the effect of these inaccuracies on the variables x_e and the costs C_e of the economical alternative or, as it is said, to test the cost equation for sensitivity. This test constitutes the third task of quantitative analysis. Once it is carried out, we can fix economically the degree of accuracy that the input or initial data should have.

Sometimes, practical considerations (the conditions of a practical mechanical process, availability of supplies, requirements to standardize plant and equipment, etc.) make it necessary to assign other than economical values to the variables. This immediately leads to a rise in the discounted costs. The greater the departure from x_e, the greater the increase in C.

It is of important practical interest to trace the character of this change, and the fourth task of technical-economic analysis is to establish how the costs C vary due to the departures of x from x_e. The procedures used for the purpose are often referred collectively to as the *test of an installation for economic stability.*

When we stated the task of finding an economical alternative, we ignored the technical constraints which are part of any technical-economic model [see Eq. (3.1)]. This is a legitimate approach where the economical value of the variable satisfies the constraints. In the plot of Fig. 3.1, this case is represented by curve *I*.

If, however, the value of x_e occurs outside the permissible limits (curve *II*), the problem must be solved so as to satisfy the constraints, that is, so as to find an optimal alternative. This constitutes the fifth task of technical-economic analysis.

In the plot of Fig. 3.1, the optimal alternative is represented by point *E*. As is seen, $C_{20}>C_{2e}$, that is, the costs of the optimal alternative exceed those of the economical alternative. This relation will also hold for problems with n variables being optimized, provided that the initial model has a single extremum. In some cases (for example, point *C* in Fig. 3.1), the optimal and economical alternatives are the same:

$$C_0=C_e$$
$$x_0=x_e$$

that is, the economical alternative is an ideal limit which is not always achievable. The fifth task can be handled by use of nonlinear integer programming and, in some cases, also by geometric programming.

It is to be noted, however, that the model, that is the cost equation and the functional constraints, Eqs. (3.16), are power polynomials (posynomials). Therefore, in order to handle the complete set of tasks of technical-economic analysis, it is advantageous to use generalizing techniques based on the theory of similarity.

The order in which the tasks defined above are to be handled and, in

fact, their number depend on the purpose of technical-economic analysis. In some cases, the consideration may well be limited to a few of them. For example, analysis for economic stability involves only the second and fourth tasks.

3.4. Use of the Theory of Similarity in Technical-Economic Analysis

Quite a number of well developed methods exist today that can be used to handle the tasks of technical and economic analysis formulated above. One such method is the method of alternative designs. However, this method is very labour and time consuming, and one usually has to limit the number of alternative designs considered. Obviously, there is nothing to guarantee that the selected alternatives include an optimal one. Also, the chance features of specific installations may mask the more general relationships. As a way out, resort has to be made to idealization and averaging, and discrete changes in parameters have to be replaced by continuous ones. This approach makes it possible to build a mathematical model and to handle optimization problems analytically, by use of differential and variational calculus and maximal programming.

The above methods are, however, applicable to complex models only when they are realized on high-speed digital computers. This is where serious difficulties are encountered. They arise not only because power-system problems are difficult to describe mathematically or because mathematical methods for their solution are not well developed yet, but also because computers have insufficient speed and limited storage capacity.

In the search for better methods to handle time and labour consuming problems of technical and economic analysis, attempts have been made to use the generalized methods borrowed from the theory of similarity. It should be remembered that analysis of complex technical and economic problems has to be conducted on the basis of incomplete information, and the problems have to be re-run many times, each time using changed variables and changed models. As a result, a wealth of numerical material is accumulated, in which it is not at all easy to establish the most typical relations, especially where the variables are many. Although problems have to be re-run many times and to the full extent, the results of the previous runs cannot be practically utilized in view of their multitude.

This explains why the need exists for generalized methods of technical and economic analysis, which would enable the input data and the results of the previous runs to be utilized far more efficiently.

The application of the theory of similarity to the tasks of technical and economic analysis, stated in Sec. 3.3, differs in several respects from its application to research and development in physics or technology. For example, it has been found that when appropriate criteria (dimensionless groups) are applied to technico-economic problems consistently, it is possible not only to extend the results of one model to a multitude of other similar models, but to carry out (as will be shown) technico-economic ana-

lysis where the numerical values of the generalized constants A are not known.

No other methods of technical and economic analysis have so far afforded this possibility.

In the analysis of optimization problems, similarity theory has revealed important technical and economic relationships in the structure of power systems, which are not unlike the physical laws of nature.

Because the use of similarity theory in technical and economic analysis is based on criteria in the form of dimensionless groups, we shall call this approach *criterion analysis*. Of course, criterion analysis is not to be set against other techniques used in technical and economic analysis. It has both merits and demerits, and it is advantageous in certain definite fields. In greater detail, this matter will be discussed later.

3.5. Dimensional Units and Per-Unit Quantities

Apart from named or dimensional units, electrical engineering widely uses dimensionless, or per-unit quantities, especially in dealing with electromagnetic and electromechanical transients in electric systems and machines. This simplifies theoretical deductions, gives them a more generalized form, and renders calculations easier to visualize in many cases. Per-unit quantities have proved useful in technical and economic analysis, too.

The *per-unit value* of a physical, economic or technico-economic quantity a is its ratio to another quantity, a_b, of the same dimensions, assumed as the reference. It is clear then that before we can derive any per-unit quantity, we should choose an appropriate reference quantity. In the calculation of short-circuit currents, for example, it is customary to build a system of reference quantities from a set of nominal or arbitrary values.

For a better understanding of the difference between dimensional and per-unit quantities, we shall consider the following example. Suppose we are to measure a voltage, V, across the terminals of a load connected to a supply line operating at a nominal voltage of $V_{\text{nom}}=220$ V. The unit of measure in this case is one volt, adopted by international agreement. The nominal voltage of the supply line is 220 times the unit of measure.

Suppose also that the actual voltage across the load terminals is measured with a voltmeter calibrated in these units of measure and that the voltmeter gives a reading of 242 scale divisions. In other words, the voltage across the load terminals is 242 times the adopted unit of voltage. Hence, in accordance with the adopted convention, we may write

$$V=242 \text{ V} \tag{3.19}$$

Alternatively, this result can be expressed as a fraction of the nominal supply voltage. For this purpose, the voltmeter indication is divided into the nominal supply voltage:

$$V_{*\,\text{nom}}=242 \text{ V}/220 \text{ V}=1.1 \tag{3.20}$$

The same result can be expressed as a fraction of an arbitrary reference voltage, say, $V_r = 2000$ V. Then,

$$V_{*r} = 242 \text{ V}/2000 \text{ V} = 0.121 \tag{3.21}$$

In the last two equations the asterisk (*) indicates that the voltage V is expressed in per-unit, while the subscripts 'nom' and 'r' designate that the unit of measure is either the nominal supply voltage or some reference voltage, respectively.

From comparison of $V_{*\text{nom}}$ and V_{*r}, it is seen that the change in the unit of measure leads to a changed numerical value of the measurand; the greater the unit of measure, the smaller the numerical value of the measurand.

It is important to note the form in which the above equations are written. Equation (3.19) is said to be in terms of named or dimensional units. The adopted unit of measure, that is 1 V, is stated on the right-hand side of the equation, because the number 242 is followed by the word *volts*. In Eqs. (3.20) and (3.21), on the other hand, the unit of measure is defined by an appropriate subscript attached to the symbol of the quantity being measured, V. This is why the right-hand sides of these equations contain only numbers, 1.1 and 0.121. Because of this, Eqs. (3.20) and (3.21) are said to be in dimensionless (per-unit) form.

However, the above division of equation forms into dimensional and dimensionless, although adopted generally, is arbitrary. For example, by analogy with Eqs. (3.20) and (3.21), we might write Eq. (3.19) in dimensionless form, too, as $V_{*v} = 242$. Then the unit of measure would be stated on the left-hand side of the equation by attaching the appropriate subscript ('v') to the measurand and using 1 volt as the reference quantity.

We might as well proceed the other way around, by writing, say, Eq. (3.20) in dimensional form, using the nominal voltage (220 V) as the unit of measure and stating the unit of measure on the right-hand side by analogy with Eq. (3.19): $V = 1.1$ of the nominal voltage (220 V).

In our further discussion, an equation will be considered to be in dimensional form if the reference unit is chosen from among those of some adopted system, such as the International System of Units (SI). If some arbitrary reference quantity be used, an equation will be considered to be in dimensionless (per-unit) form.

Generally speaking, an arbitrary reference may be adopted for any quantity included in a technical-economic model, independently of the reference units chosen for other quantities. For example, in the model described by Eq. (3.7) the reference quantities may be C_r, A_{0r}, A_{1r}, A_{2r} and F_r. No constraints need be imposed on the choice of numerical values for the reference quantities, except the natural requirement stemming from the technical or economic nature of the problem, namely that the quantities should be real, non-zero or finite.

If reference quantities are to be set up in a system, one more condition should be satisfied:

$$C_r = A_{0r} + A_{1r}F_r + A_{2r}F_r^{-1} \tag{3.22}$$

Let us define similarity criteria for the above equation which relates all the variables and constants of interest. This can be done on the basis of the integral analog method, by dividing Eq. (3.22) through by C_r:

$$1 = A_{0r}/C_r + A_{1r}F_r/C_r + A_{2r}F_r^{-1}/C_r \tag{3.23}$$

Each term on the right-hand side of Eq. (3.23) is a similarity criterion by definition,

$$A_{0r}/C_r = \pi_{0r}, \quad A_{1r}F_r/C_r = \pi_{1r}, \quad A_{2r}F_r^{-1}/C_r = \pi_{2r} \tag{3.24}$$

In per-unit form, the present values of the quantities, the values other than the reference ones, may be expressed as:

$$C_{*r} = C/C_r, \quad A_{0*r} = A_0/A_{0r}, \quad A_{1*r} = A_1/A_{1r}$$
$$A_{2*r} = A_2/A_{2r}, \quad F_{*r} = F/F_r$$

Substituting the per-unit and reference values for the respective quantities, C, A_0, A_1, A_2 and F, in Eq. (3.7), yields

$$C_{*r}C_r = A_{0*r}A_{0r} + A_{1*r}A_{1r}F_{*r}F_r + A_{2*r}A_{2r}/(F_{*r}F_r)$$

Dividing the above equation through by C_r and noting Eq. (3.24), we get

$$C_{*r} = \pi_{0r}A_{0*r} + \pi_{1r}A_{1*r}F_{*r} + \pi_{2r}A_{2*r}/F_{*r} \tag{3.25}$$

The above expression gives little to go by in technical and economic analysis. For one thing, it is more unwieldy than the original equation, (3.7). For another, it does not reduce the quantity of data needed.

However, we may impose further constraints on the numerical values of the reference quantities. If we assume that

$$A_{0r} = A_0, \quad A_{1r} = A_1, \quad A_{2r} = A_2$$

then

$$A_{0*} = A_{1*} = A_{2*} = 1$$

The above condition is quite legitimate, because the constants A_i may take on any value between zero and infinity, including those stated above. In such a case,

$$C_{*r} = \pi_{0r} + \pi_{1r}F_{*r} + \pi_{2r}F_{*r}^{-1} \tag{3.26}$$

Thus, we have simplified Eq. (3.25). A very important feature about Eq. (3.26) is that all of its terms are in per-unit, or dimensionless. This example shows that the original discounted cost equation, Eq. (3.7), can be converted to dimensionless form. Through a proper selection of values for the reference quantities, the form can be greatly simplified [compare Eqs. (3.25) and (3.26)].

3.6. Economical Alternative

A distinction of criterion analysis applied to technico-economic studies is that the reference quantities are the economical values of variables and costs.

Using Eq. (3.5) as an example, we shall show how the set of reference quantities can be selected. For convenience, we designate

$$C_0=(p_{st}+a/100)K_0'l$$
$$a_F=(p_{st}+a/100)K_0''$$

and write the line cost equation as

$$C_{line}=C_0+a_Fl F+3I_{\max}^2 b\tau\varrho l/F \tag{3.27}$$

From the above equation it follows that an increase (decrease) in the cross-sectional area of conductors will increase (decrease) the second term of the model and, at the same time, reduce (increase) the third term. Thus, in this model there is a competition between the second and third terms as regards cross-sectional area. No other quantity in the model causes such a competition. With the other quantities held fixed, for each value of F there is a definite value of costs. By varying F, we can trace the relationship between the constituent cost items and the total cost. In doing so, we will note that there is a cross-sectional area for which the costs will be a minimum (Fig. 3.2). This minimum can be evaluated by equating to zero the first derivative of the costs with respect to cross-sectional area:

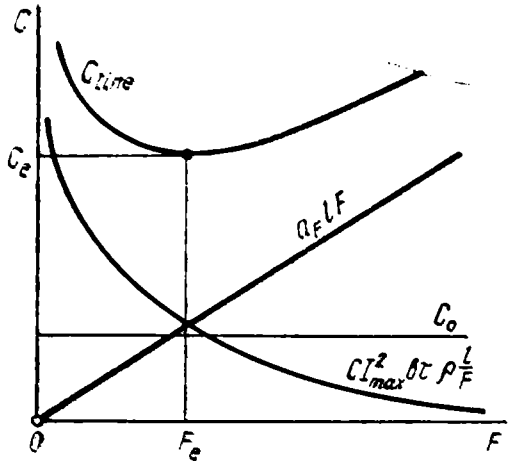


Fig. 3.2

$$dC_{line}/dF=0$$

whence

$$F_e=I_{\max}\sqrt{3b\tau\varrho/a_F} \tag{3.28}$$

The quantities entering Eq. (3.27) may only take positive values, and so the second derivative is

$$d^2C_{line,e}/DF^2>0$$

It follows that the costs C will be a minimum at F_e, as suggested by the plot. Substituting this value in Eq. (3.27) yields the minimum costs of the line

$$C_{line,e}=C_0+2I_{\max}l\sqrt{3b\tau a_F\varrho} \tag{3.29}$$

It is seen from Eq. (3.28) that the economical cross-sectional area of the line conductors is independent of the fixed cost item C_0. Therefore, in the selection of F_e it may be neglected, and consideration may be given to only

the variable cost term which is dependent on changes in the objective variable, F:

$$C=C_{line}-C_0=a_F lF+3I_{max}^2 b\tau\varrho l/F \tag{3.30}$$

If we move C_0 to the left-hand side, the minimum value of the variable cost term can be found from Eq. (3.29):

$$C_e=C_{line,e}-C_0=2I_{max}l\sqrt{3b\tau a_F\varrho} \tag{3.31}$$

This value will be referred to as *economical.* As is noted in Sec. 3.3, it is one of the principal tasks of technical and economic analysis to determine an economical alternative. The values of F_e and C_e found by Eqs. (3.28) and (3.31) are used as the reference quantities in criterion analysis.

3.7. Technical and Economic Proportioning

Once an economical alternative has been selected, the second problem of technical and economic analysis, that of economical proportioning, can be tackled. As before, we shall use the example of Sec. 3.6.

The line cost equation, Eq. (3.30), has two terms. The first term represents the variable cost proportional to conductor cross section, and the second term the costs proportional to the energy lost in the line, which vary inversely as conductor cross section. To begin with, we shall convert the model to dimensionless form by dividing all equations into the sum of variable costs:

$$1=a_F lF/C+3I_{max}^2 b\tau\varrho l/CF$$

The dimensionless groups

$$\begin{aligned} a_F lF/C&=\pi_1 \\ 3I_{max}^2 b\tau\varrho l/CF&=\pi_2 \end{aligned} \tag{3.32}$$

are, in fact, the similarity criteria found by the integral analog method [4]. In our case, they are simple to interpret: they show the share of the cost proportional to conductor cross section, π_1, and of the cost proportional to the energy lost in the line, π_2. The similarity criteria are expressed in per-unit with the total costs, C, chosen as reference, such that

$$\pi_1+\pi_2=1 \tag{3.33}$$

Written in general form, this criterion equation describes a multitude of alternatives to Eq. (3.30). Consider its features.

With Eq. (3.30), we can identify a particular alternative design of the line by assigning numerical values to all of its quantities. No single alternative design can be identified from Eq. (3.33). By assigning numerical values to the similarity criteria we can only specify a particular alternative of the generalized equation, which will in turn describe a set of alternative line designs, and not a single alternative. A distinction of this generalized alternative is that it covers a multitude of individual alternatives to the

design equation and that the similarity criteria in it remain numerically constant in all of these individual alternatives

$$\pi_1 = \text{idem}, \quad \pi_2 = \text{idem}$$

By the first theorem of similarity theory, such alternatives are similar.

Changing the numerical values of the similarity criteria leads to a new set of similar alternatives. Thus, the numerical values of the similarity criteria form the basis for a quantitative assessment of similar alternatives.

Consider how an economical alternative can be found by use of the criterion equation (3.33). By assigning numerical values to the similarity criteria π_1 and π_2, we can select from all the likely alternatives of Eq. (3.30) only subsets of similar alternatives which may widely differ in the value of all the quantities included in the dimensionless groups, Eqs. (3.32). Clearly the generalized equation, Eq. (3.33), must include economical alternatives, too. Therefore, the selection of economical alternatives in generalized form reduces to determining the numerical values of the similarity criteria, π_{ie}, associated with economical alternatives.

In other words, when use is made of criterion equation (3.33), the search for an economical alternative is to establish an economical break-down between the terms of the cost equation, Eq. (3.30), that is, to determine economical itemized costs.

Let us see how an alternative design with economically well-proportioned cost items can be distinguished from any other. To do this, we shall first determine the numerical values of the similarity criteria, Eqs. (3.32), for an economical alternative by substituting appropriate expressions for F_e and C_e:

$$\pi_{1e} = 1/2, \quad \pi_{2e} = 1/2 \tag{3.34}$$

Thus, for economical alternative designs of the power transmission line, the variable cost items proportional to the conductor cross section, π_{1e}, should be equal to those proportional to the energy lost in the line, π_{2e}. In the literature, this is known as *Kelvin's law.*

As already noted, assignment of numerical values to the similarity criteria identifies a generalized alternative design covering a set of similar individual alternative designs of the line. By assigning value 1/2 to the similarity criteria, we thus isolate economical alternatives from the complete set of alternatives. Hence, we may conclude that *the economical alternative designs of a power transmission line are similar**.

From Eq. (3.33) it follows that the numerical value of one similarity criterion is uniquely defined by that of the other. Hence, the following ratio can be used in order to find the similarity criteria:

$$\pi_2/\pi_1 = k_F \tag{3.35}$$

which will be called *Kelvin's secondary similarity ratio.*

* The similarity of economical alternatives for the *n*-ary case is discussed in Sec. 3-12.

For the economical alternative designs of a transmission line this similarity ratio is unity:

$$k_F^e=\pi_{2e}/\pi_{1e}=1 \tag{3.36}$$

The last relation is a mathematical description of a feature possessed only by alternative line designs with economically balanced or proportioned cost items.

3.8. Similarity Criteria for Economical Alternatives

In the general case, the numerical values of economical similarity criteria can be found by Eqs. (3.18). This requires knowledge of the value of not only the variable being optimized, x_e, and of the minimum cost, C_e, but also of all the constants A_i entering the equations.

We set out to show that in some cases the economical values of similarity criteria depend solely on the exponents α_{ij} of the variables being optimized. Let the design equation be of the form

$$C=A_1x^{\alpha_1}+A_2x^{-\alpha_2} \tag{3.37}$$

which is a generalization of the transmission-line cost equation, Eq. (3.30). For an economical alternative, this equation may be written as

$$C_e=A_1x_e^{\alpha_1}+A_2x_e^{-\alpha_2}$$

The economical similarity criteria can be found from the ratios:

$$\pi_{1e}=(A_1/C_e)x_e^{\alpha_1}, \quad \pi_{2e}=(A_2/C_e)x_e^{-\alpha_2} \tag{3.38}$$

and their values are such that their sum is unity

$$\pi_{1e}+\pi_{2e}=1 \tag{3.39}$$

To determine the economical value of the variable, x_e, we equate to zero the first cost derivative, Eq. (3.37), with respect to the variable x:

$$dC/dx=\alpha_1A_1x_e^{\alpha_1-1}-\alpha_2A_2x_e^{-\alpha_2-1}=0$$

Substituting Eq. (3.38) for A_1 and A_2 in the above equation gives:

$$\alpha_1\pi_{1e}C_e/x_e-\alpha_2\pi_{2e}C_e/x_e=0$$

Since $C_e/x_e\neq 0$, then

$$\alpha_1\pi_{1e}-\alpha_2\pi_{2e}=0 \tag{3.40}$$

By solving the above equation simultaneously with Eq. (3.39) we get

$$\begin{aligned}\pi_{1e}&=\alpha_2/(\alpha_1+\alpha_2)\\ \pi_{2e}&=\alpha_1/(\alpha_1+\alpha_2)\end{aligned} \tag{3.41}$$

As is seen, the numerical values of the similarity criteria for the economical alternatives of Eq. (3.37) are decided by the exponents α_1 and α_2 of the objective variable (one being optimized), as has been postulated.

Example 3.1. Evaluate the similarity criteria for the economical alternative of the equation

$$C=A_1x^2+A_2x^{-5}$$

The exponents in the above equation are:

$$\alpha_1=2, \qquad \alpha_2=5$$

By Eq. (3.41), the similarity criteria for the economical alternatives will be

$$\pi_{1e}=5/(2+5)=5/7$$
$$\pi_{2e}=2/(2+5)=2/7$$

3.9. Technico-Economic Stability

In Sec. 3.6, the conductor cross-sectional area was selected without regard for the technical restrictions arising from the fact that the conductors are only available in discrete sizes, that the voltage drop should be kept within specified limits, etc. On the other hand, it is a well-known fact that compliance with technical restrictions may raise the cost of a line. It is important to know by how much the actual cost may then depart from the economical value. This problem can advantageously be solved, using a criterion model of the line and adopting the economical alternative as the reference.

We shall illustrate how a criterion model can be set up, using Eqs. (3.37) as an example. Since we are concerned with the departures of the cost C and the variable x from their economical values, we may write

$$C_*=C/C_e$$
$$x_*=x/x_e$$

Substituting the above expressions for C and x in Eqs. (3.37) gives

$$C_*=\frac{A_1x_e^{\alpha_1}}{C_e}x_*^{\alpha_1}+\frac{A_2x_e^{-\alpha_2}}{C_e}x_*^{-\alpha_2}$$

or, noting Eqs. (3.38) and (3.41),

$$C_*=\frac{\alpha_2}{\alpha_1+\alpha_2}x_*^{\alpha_1}+\frac{\alpha_1}{\alpha_1+\alpha_2}x_*^{-\alpha_2} \tag{3.42}$$

Using this criterion equation, we can trace how departures in the variable x from its economical value x_e affect the relative increase in cost C_*.

Notably, when $\alpha_1=\alpha_2=1$, Eq. (3.42) takes a simpler form

$$C_*=0.5(F_*+F_*^{-1}) \tag{3.43}$$

A plot of this equation appears in Fig. 3.3*a*. The same equation is plotted in Fig. 3.3*b* so that the economical values of the variable, $F_*=1$, and of the cost, $C_*=1$, occur at the origin of coordinates. This coordinate system is more convenient in criterion analysis and will be used frequently in the subsequent discussion.

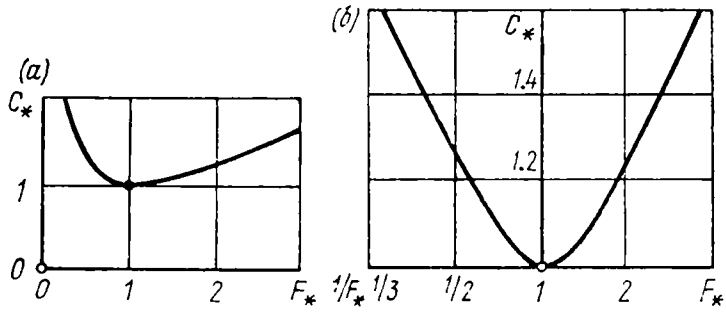


Fig. 3.3

In generalized form, criterion equation (3.43) describes all alternative designs of the power transmission line and makes it possible to determine directly the relative increase in cost, C_*, due to a relative departure of the variable from its economical value.

As follows from Eq. (3.42), the form of the curve $C_*=f(x_*)$ is decided by the exponents of the variable being optimized, that is, by the values of α_1 and α_2. According to the values of these exponents, equations of the form of Eq. (3.37) may be divided into economically stable and economically unstable.

If large departures of the variables from x_e lead to a small increase in the discounted cost, the equation under consideration is *economically stable*. Conversely, the equation is *economically unstable*, when a small departure of x from x_e brings about a considerable increase in the discounted cost. In the plot of Fig. 3.4, curve *II* represents an equation which is less stable than that represented by curve *I*.

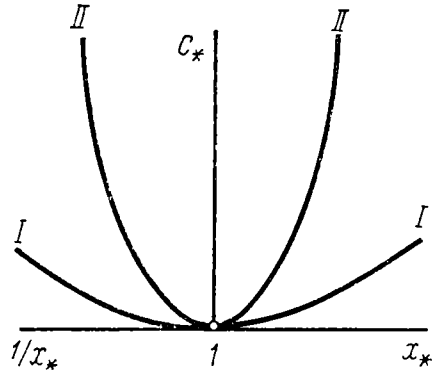


Fig. 3.4

Although stability is a relative term, it can be treated quantitatively as well. We shall agree to call the discounted cost equation stable, if a 5-per cent departure of x from x_e leads to an increase of not more than $\delta=1.05$ in the cost. Sometimes a greater or a smaller value may be assigned to δ, depending on the nature of the problem at hand and some other factors.

The design discounted cost equation should preferably be tested for economical stability by use of criterion relations analogous to Eqs. (3.42), as they make knowledge of the constants A_i unnecessary.

3.10. Analysis for Technico-Economic Sensitivity

As a rule, power-system problems are analysed for technical and economic indicators, using known input data. It is a frequent occurrence, however, that input data are guessed or assumed. This is particularly true when

a novel type of plant or equipment is being designed, for which no prior art of practical value exists. Since these input data are inaccurate, an alternative design found by rigorous mathematical methods would have a dubious value, because every improvement in the input data would lead to a different alternative. This is why the model under consideration must be tested for *technico-economic sensitivity*. In other words, it is essential to see how inaccuracies in input or design data may affect the selection of optimal alternatives. Also, it is important to know how this selection can be affected by each quantity involved. Once the two factors are established, we can define the degree of accuracy that the input or design data should have. If a considerable inaccuracy in any data item leads, in the final analysis, to an insignificant change in the optimal alternative, there is no sense in looking for the precise value of the quantity involved. In contrast, if small changes in some data might lead to considerable changes in the optimal alternative, the relevant data must be specified with maximum accuracy.

Models should be tested for technico-economic sensitivity, preferably using equations similar to Eqs. (3.28) and (3.31).

Suppose that for some initial value of the variables adopted as the reference ones, the economical values of the variable being optimized and of the cost (for $l=1$ km) are as follows:

$$\left.\begin{aligned} F_{e,r} &= (3I^2_{\max r} b_r \tau_r \varrho_r / a_{Fr})^{1/2} \\ C_{e,r} &= 2(3I^2_{\max r} b_r \tau_r \varrho_r a_{Fr})^{1/2} \end{aligned}\right\} \tag{3.44}$$

For other input data, the values of F_e and C_e can be found by Eqs. (3.28) and (3.31). Then, dividing the equations thus obtained by those of (3.44) gives

$$\left.\begin{aligned} F_{*e} &= (I^2_{\max *} b_* \tau_* \varrho_* / a_{F*})^{1/2} \\ C_{*e} &= (I^2_{\max *} b_* \tau_* \varrho_* a_{F*})^{1/2} \end{aligned}\right\} \tag{3.45}$$

Using the last system of equations, it is possible to test the model for sensitivity towards any quantity entering the initial equation, or their combinations. In doing so, we need not find an economical or any other alternative.

As an example, we shall determine the error in F_e when all input or design data in Eq. (3.27) are accurate to within $\pm 5\%$. From Eq. (3.45) it follows that F_e will be a maximum when

$$I_{\max *} = b_* = \tau_* = \varrho_* = 1.05$$

and

$$a_{F*} = 0.95$$

Substituting these values in the expression for F_{*e} gives

$$F^+_{*e} = (1.05^2 \times 1.05 \times 1.05 \times 1.05/0.95)^{1/2} = 1.16$$

F_e will be a minimum when

$$I_{\max *} = b_* = \tau_* = \varrho_* = 0.95$$

and

$$a_{F*}=1.05$$

In this case,

$$F_{*e}^{-}=(0.95^2\times0.95\times0.95\times0.95/1.05)^{1/2}=0.86$$

As is seen, for the design data accurate within $\pm5\%$, the per cent error in F_e is in the range from $+16\%$ to -14%.

Conductors for low-voltage (6 to 10 kV) transmission lines have cross sections graded so that the ratio of two adjacent sizes is about 1.4. Assuming that the input data have the accuracy defined above, the maximum error in the selection of conductor size for these lines should not exceed one size gradation.

Similar calculations show that the likely error in C_e in such a case will be in the range from $+27\%$ to -18%. Since the variable cost item of C_e is ordinarily 20 to 30% of the total cost of the line (including the fixed costs), the maximum likely error in the computed cost of the line, with design data accurate within $\pm5\%$, will not exceed 5 to 10%. Conductor sizes need not be specified with greater accuracy because it corresponds to the existing gradations of conductor sizes.

3.11. Range of a Variable

In the economic analysis of an installation, machine or apparatus, especially at the design stage, several discrete values, making up a nominal range of values, have to be considered for each variable, such as cross section, voltage, power, etc. The crux of the problem lies in the fact that in most cases the design data can be found only approximately. Therefore, in handling this problem, it is advisable to use relative variable ranges. This method will be explained, using the choice of a range of discrete sizes for conductors as an example.

Suppose we are to find the reasonable number of steps into which the range of change ΔF in the variable F is to be divided where

$$\Delta F=F_{\max}-F_{\min}$$

If conductors were commercially available in only one cross section, the discounted cost would be a minimum only when the proposed design used this available cross section. In all other cases, there would be a departure from the minimum likely cost (Fig. 3.5*a*). If conductors were made available in, say, three sizes (Fig. 3.5*b*), with other conditions being equal, the relative increase in the cost would be smaller. However, too wide a range of standard conductor sizes is not a necessity from the view-point of design accuracy, because design data are in error always. This error must be taken into account in determining the spacing between adjacent standard conductor sizes.

It follows from Fig. 3.5 that the ith standard conductor size (cross-sectional area) should lie somewhere between a minimum value, $F_{i\min}$, and

a maximum value, $F_{i\max}$, which respectively put the lower and upper bound on the ith interval of the variable F:

$$\Delta F_i = F_{i\max} - F_{i\min} \tag{3.46}$$

It is natural to assume that the expectation of the required conductor sizes of the ith interval should lie within this range and be equal to F_{ie}. We adopt this value as the reference one and write the variable-range equation, Eq. (3.46), in per-unit form:

$$\begin{aligned}\Delta F_{i*} &= \Delta F_i / F_{ie} \\ &= F_{i\max}/F_{ie} - F_{i\min}/F_{ie} \\ &= F_{i*\max} - F_{i*\min}\end{aligned}$$

Thus, to determine the relative (per-unit) range ΔF_{i*}, we need to know the per-unit value of the upper bound $F_{i*\max}$ and of the lower bound $F_{i*\min}$ of the range. This can be done, using criterion equation (3.43).

If we assume that within the range considered the model is economically stable, $C_* \leqslant \delta$, then, noting Eq. (3.43), the last inequality may be written as

$$2\delta \geqslant (F_* + F_*^{-1})$$

Solving it for F_*, we obtain for the limiting case

$$F_* = \delta \pm \sqrt{\delta^2 - 1}$$

Then we find the upper and the lower bounds on the relative (per-unit) conductor size range corresponding to the adopted accuracy of costing:

$$F_{*\max} = \delta + \sqrt{\delta^2 - 1}$$

$$F_{*\min} = \delta - \sqrt{\delta^2 - 1}$$

and, as a consequence, the relative (per-unit) range itself:

$$\Delta F_* = F_{*\max} - F_{*\min} = 2\sqrt{\delta^2 - 1} \tag{3.47}$$

which defines the spacing between conductor sizes from economical

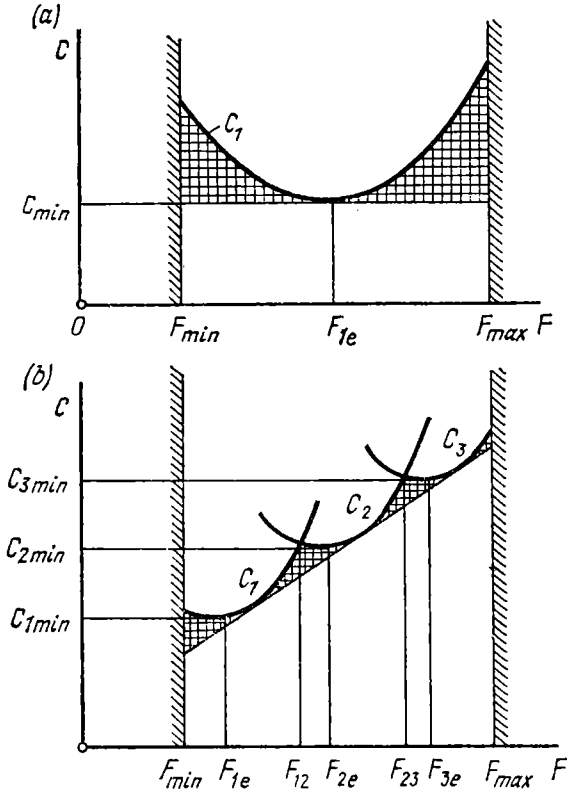


Fig. 3.5

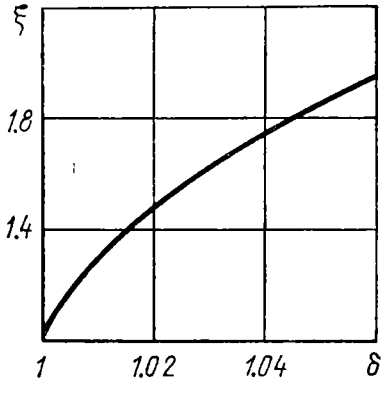


Fig. 3.6

considerations and in view of the required accuracy in design. In practice, it is more convenient to replace the relative (per-unit) range, Eq. (3.47), by the upper to lower bound ratio:

$$\xi = F_{\max}/F_{\min} = F_{*\max}/F_{*\min} = [\delta + (\delta^2 - 1)^{1/2}]^2 \tag{3.48}$$

A plot of this equation appears in Fig. 3.6. It relates the bound ratio to the costing accuracy required. For example, at $\delta = 1.02$, $\xi = 1.5$; at $\delta = 1.05$, $\xi = 1.85$. It should be remembered that these figures have been obtained from an analysis of an approximate line cost equation. If we include the fixed term, the figures will be different.

3.12. Canonical Form of the Discounted Cost Equation

In the previous sections we have analysed problems involving one objective variable. We have seen that criterion relations (dimensionless groups) make it possible to carry out technico-economic analysis even when the numerical values of the design data, the constants A_i, are not known. No other existing methods afford this possibility. Criterion analysis brings out the relationships that hold for all economical alternative designs, irrespective of the specific values of design data, and that may be used to allocate resources among several projects or expansion schemes. Also, this form of analysis offers fairly simple means by which designs can be tested for technico-economical stability and sensitivity and the necessary and sufficient limits of accuracy can be defined for both the design data and the economical alternative design.

As already noted, however, the above examples are simple and have been chosen only to demonstrate the subject-matter. Because of this, they cannot give full justice to the theory of similarity as a tool in the selection of the most economical alternative. The sections that follow set forth criterion analysis in its application to the more complex technico-economic problems.

Let the design discounted cost equation be a generalized polynomial* of the form

$$C = \sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_j^{\alpha_{ij}} \tag{3.49}$$

To begin with, we determine the analytical relationships required to find economical values for the variables, x_{je}, and the cost, C_e. For the alternatives of interest, Eq. (3.49) takes the form

$$C_e = \sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_{je}^{\alpha_{ij}} \tag{3.50}$$

* Sometimes, generalized polynomials of this form are called generalized posynomials.

It can be converted to dimensionless form by dividing it into C_e:

$$1=\sum_{i=1}^{m}\frac{A_i}{C_e}\prod_{j=1}^{n}x_{je}^{\alpha_{ij}} \tag{3.51}$$

From the theory of similarity it is known that this operation yields the quantities

$$\frac{A_i}{C_e}\prod_{j=1}^{n}x_{je}^{\alpha_{ij}}=\pi_{ie} \tag{3.52}$$

which are similarity criteria. Since C_e and x_{je} are defined for economical alternatives, the π_{ie}'s should likewise be termed economical similarity criteria. They show how much each term of the original equation, Eq. (3.50), contributes to the total costs of the sought alternative design. That is, they serve as a fractional measure of each cost item. From Eqs. (3.51) and (3.52) it follows that the sum of the similarity criteria is unity:

$$\sum_{i=1}^{m}\pi_{ie}=1 \tag{3.53}$$

If we determine the A_i's from Eq. (3.52) and substitute the result in Eq. (3.49), we obtain

$$C=\sum_{i=1}^{m}C_e\pi_{ie}\prod_{j=1}^{n}x_j^{\alpha_{ij}}\Big/\prod_{j=1}^{n}x_{je}^{\alpha_{ij}}$$

or, in per-unit form,

$$C_*=\sum_{i=1}^{m}\pi_{ie}\prod_{j=1}^{n}x_{j*}^{\alpha_{ij}} \tag{3.54}$$

Comparison of Eqs. (3.49) and (3.54) shows that in order to write a criterion equation, it is necessary to replace the constants in the design equation by the appropriate similarity criteria, and the variables and costs by per-unit quantities.

Equation (3.52) makes it possible to evaluate the economical similarity criteria π_{ie} when the constants A_i and the values of x_{ie} and C_e are known. If the π_{ie}'s were known in advance, the same equations could be used to determine the economical values of the variables and costs.

As a proof, we write Eq. (3.52) as

$$\frac{1}{C_e}\prod_{j=1}^{n}x_{je}^{\alpha_{ij}}=\frac{\pi_{ie}}{A_i} \tag{3.55}$$

and take logarithms

$$\left.\begin{aligned}\alpha_{11}\ln x_{1e}+\ldots+\alpha_{1n}\ln x_{ne}-\ln C_e&=\ln\pi_{1e}-\ln A_1\\ \alpha_{m1}\ln x_e+\ldots+\alpha_{mn}\ln x_{ne}-\ln C_e&=\ln\pi_{me}-\ln A_m\end{aligned}\right\} \tag{3.56}$$

Designating

$$\ln x_{ie}=Y_{ie},\qquad \ln C_e=Z_e,\quad\text{and}\quad \ln\pi_{ie}-\ln A_i=K_i$$

we can re-write Eqs. (3.56) as

$$\alpha_{11} Y_{1e} + \ldots + \alpha_{1n} Y_{ne} - Z_e = K_1$$
$$\alpha_{m1} Y_{1e} + \ldots + \alpha_{mn} Y_{ne} - Z_e = K_m$$

or, in contracted matrix form

$$\alpha \mathbf{Y}_e = \mathbf{K}_e \tag{3.57}$$

where the exponent matrix α, the economical-variable column matrix $\mathbf{Y}_e$ and the constant matrix $\mathbf{K}_e$ have the following form

$$\alpha = \left\| \begin{matrix} \alpha_{11} \ldots \alpha_{1n} - 1 \\ \ldots\ldots\ldots\ldots \\ \alpha_{m1} \ldots \alpha_{mn} - 1 \end{matrix} \right\|, \quad \mathbf{Y}_e = \left\| \begin{matrix} Y_{1e} \\ \ldots \\ Y_{ne} \\ Z_e \end{matrix} \right\|, \quad \mathbf{K}_e = \left\| \begin{matrix} K_1 \\ \ldots \\ K_m \end{matrix} \right\| \tag{3.58}$$

If $m = n+1$ and the exponent (α) matrix (the dimension matrix) is non-singular, then we may use Eq. (3.57) to determine the column matrix

$$\mathbf{Y}_e = \alpha^{-1} \mathbf{K}_e \tag{3.59}$$

By taking an anti-logarithm of the above equation, we find the economical alternative

$$\exp Y_{je} = \prod_{i+1}^{n+1} \left(\frac{\pi_{ie}}{A_i} \right)^{\Delta_{ji}/\Delta}, \quad j = \left| \begin{matrix} n+1 \\ 1 \end{matrix} \right. \tag{3.60}$$

where Δ_{ji}/Δ are the elements of the inverse (α^{-1}) matrix.

From Eq. (3.60), a particular case arises

$$\exp Y_{(n+1)e} = C_e = \prod_{i=1}^{n+1} \left(\frac{\pi_{ie}}{A_i} \right)^{\Delta_{n+1,\, i}/\Delta} \tag{3.61}$$

$$\exp Y_{je} = x_{je} = \prod_{i=1}^{n+1} \left(\frac{\pi_{ie}}{A_i} \right)^{\Delta_{ji}/\Delta}, \quad j = \left| \begin{matrix} n \\ 1 \end{matrix} \right. \tag{3.62}$$

Thus, if the design data, or constants, A_i and the similarity criteria π_{ie} are specified, an economical alternative can be defined only when the exponent matrix of the variables being optimized is non-singular. Then both the variables x_e and the costs C_e are evaluated by equations identical in form. This is convenient from a computational point of view in comparison with other methods used to identify an economical alternative.

The proposed method, however, requires knowledge of the numerical values of the similarity criteria, which are generally found by Eqs. (3.52). Thus, we have found ourselves in a kind of vicious circle: in order to determine Y_{je}, we must know π_{ie}; in order to evaluate π_{ie}, Y_{je} must be known. To break the vicious circle, we should devise a means for determining the similarity criteria π_{ie} in some other way, without having to compute Y_{je}. Such a means applicable to problems involving a single variable

has been examined in Sec. 3.8, where it has been shown that in some cases similarity criteria π_{ie} are dependent solely on the exponents of the variables being optimized [see Eq. (3.41)]. We shall now prove that this possibility exists for many-variable problems as well.

To begin with, we recall that the first derivatives of Eq. (3.49) with respect to the variables of the economical alternative are zero

$$\frac{\partial C}{\partial x_s}=\sum_{i=1}^{m}\frac{\alpha_{is}A_i}{x_{se}}\prod_{j=1}^{n}x_{je}^{\alpha_{ij}}=0, \qquad s=\overline{|}\begin{matrix}n\\1\end{matrix} \tag{3.63}$$

Since, in accordance with Eq. (3.52) for such alternatives,

$$A_i\prod_{j=1}^{n}x_{je}^{\alpha_{ij}}=\pi_{ie}C_e$$

Eqs. (3.63) may be re-written as

$$\frac{C_e}{x_{se}}\sum_{i=1}^{m}\alpha_{is}\pi_{ie}=0, \qquad s=\overline{|}\begin{matrix}n\\1\end{matrix} \tag{3.64}$$

According to the statement of a technico-economic problem,

$$C_e/x_{se}\neq 0$$

and so this term can be dropped from Eq. (3.64):

$$\sum_{i=1}^{m}\alpha_{is}\pi_{ie}=0, \qquad s=\overline{|}\begin{matrix}n\\1\end{matrix} \tag{3.65}$$

that is, in the case of problems in many variables the similarity criteria depend on the exponents of the variables being optimized.

Being homogeneous, Eqs. (3.65) are always simultaneous, but they enable the similarity criteria to be evaluated only for $m=n$. In such a case, the solution would be unique and trivial: $\pi_{ie}=0$, but it would have no technico-economic sense whatsoever.

In the general case, the simultaneous equations, Eqs. (3.65), will have a non-trivial (non-zero) solution only when the rank r of Eqs. (3.65) is smaller than the number of unknowns, that is, when the number of variables being optimized, n, is less than the number of cost terms, m, in the original model, Eq. (3.49). But again Eqs. (3.65) make it possible only to characterize r principal similarity criteria in terms of $(m-r)$ additional similarity criteria.

To resolve this difficulty, in addition to the orthogonality condition arising from the manner in which the extremum of several variables can be found, the constraints imposed on the similarity criteria should include the normality condition Eq. (3.53), which is a natural corollary where similarity numbers are determined by the integral analog method.

The inhomogeneous equations (3.53) and (3.65) can be re-written in contracted matrix form as

$$\alpha^T\pi_e=\beta \tag{3.66}$$

where

$$\alpha^T = \begin{Vmatrix} \alpha_{11} \dots \alpha_{m1} \\ \alpha_{1n} \dots \alpha_{mn} \\ -1 \dots -1 \end{Vmatrix}, \qquad \pi_e = \begin{Vmatrix} \pi_{1e} \\ \vdots \\ \pi_{me} \end{Vmatrix}, \qquad \beta = \begin{Vmatrix} 0 \\ \vdots \\ 0 \\ -1 \end{Vmatrix} \tag{3.67}$$

It is to be noted that the matrix α^T is a transpose of the matrix α which has earlier been used to determine an economical alternative [see Eq. (3.58)].

If the determinants of the matrix α and, as a consequence, of the matrix α^T are non-zero ($\Delta \neq 0$), the solution of Eqs. (3.66) is existent and unique:

$$\pi_e = \| \alpha^T \|^{-1} \beta \tag{3.68}$$

In this case, the similarity criteria of economical alternatives depend solely on the exponents of the variables being optimized. This case is of important practical interest; it is called *canonical*, and the original model, Eq. (3.49), for which the matrix α is non-singular is called a *canonical model* or a *model in canonical form.*

3.13. Examples of Analysis of Canonical Equations

Criterion analysis of canonical equations will be demonstrated, using several typical problems as examples.

Example 3.2. In design practice, it is often required to investigate the economical advisability of choosing a particular conductor cross-sectional area, F, and voltage, V, for a power transmission line.

Experience in the design of lines for voltages up to 110 kV shows that the above problem can be handled by use of the following dependence of discounted costs on F and V:

$$C_{line} = C_0 + A_V V + A_F F + A_s/(V^2 F) \tag{3.69}$$

In the above equation, the first term represents the fixed item of the line cost, independent of either voltage or cross section; the second term represents the costs proportional to the line voltage; the third, to the conductor cross-sectional area; and the fourth, to the energy lost in the line.

Let the line voltage V and the conductor cross-sectional area F be the variables to be optimized; in the model, Eq. (3.69), they give rise to com petition effects.

If we consider only the variable terms of the discounted costs we shall get the following design equation:

$$C = C_{line} - C_0 = A_V V + A_F F + A_s/(V^2 F) \tag{3.70}$$

which can be analysed by use of known criteria (see Sec. 3.12). To begin with, we shall establish analytical relations of the form of Eq. (3.60) to determine economical alternatives.

By analogy with Eq. (3.52), the similarity criteria of economical alternatives may be written as

$$\pi_{1e}=A_V V_e/C_e, \quad \pi_{2e}=A_F F_e/C_e, \quad \pi_{3e}=A_s/V_e^2 F_e C_e \tag{3.71}$$

Next we take the logarithms of the equations and arrange them in a tableau (array) where the independent variables $\ln V_e$, $\ln F_e$ and $\ln C_e$ occupy columns, and the dependent variables $\ln \pi_1/A_V$, $\ln \pi_2/A_F$ and $\ln \pi_3/A_s$ occupy rows:

	$\ln V_e$	$\ln F_e$	$\ln C_e$
$\ln \pi_1/A_V$	1	0	−1
$\ln \pi_2/A_F$	0	1	−1
$\ln \pi_3/A_s$	−2	−1	−1

(3.72)

The numbers inside the box form a dimension matrix. In order to solve Eqs. (3.71) for V_e, F_e and C_e, it is necessary to invert the matrix α. This is done by interchanging the dependent and independent variables in tableau (3.72). The inverse tableau is:

	$\ln \pi_1/A_V$	$\ln \pi_2/A_F$	$\ln \pi_3/A_S$
$\ln V_e$	1/2	−1/4	−1/4
$\ln F_e$	−1/2	3/4	−1/4
$\ln C_e$	−1/2	−1/4	−1/4

(3.73)

where the inverse dimension matrix α^{-1} is enclosed in the box.

We can write matrix equation (3.73) at length as follows:

$$\ln V_e=\frac{1}{2}\ln\frac{\pi_{1e}}{A_V}-\frac{1}{4}\ln\frac{\pi_{2e}}{A_F}-\frac{1}{4}\ln\frac{\pi_{3e}}{A_s}$$

$$\ln F_e=-\frac{1}{2}\ln\frac{\pi_{1e}}{A_V}+\frac{3}{4}\ln\frac{\pi_{2e}}{A_F}-\frac{1}{4}\ln\frac{\pi_{3e}}{A_s}$$

$$\ln C_e=-\frac{1}{2}\ln\frac{\pi_{1e}}{A_V}-\frac{1}{4}\ln\frac{\pi_{2e}}{A_F}-\frac{1}{4}\ln\frac{\pi_{3e}}{A_s}$$

and take the respective anti-logarithms

$$\left.\begin{aligned} V_e&=\left(\frac{\pi_{1e}}{A_V}\right)^{1/2}\left(\frac{A_F}{\pi_{2e}}\right)^{1/4}\left(\frac{A_s}{\pi_{3e}}\right)^{1/4}\\ F_e&=\left(\frac{A_V}{\pi_{1e}}\right)^{1/2}\left(\frac{\pi_{2e}}{A_F}\right)^{3/4}\left(\frac{A_s}{\pi_{3e}}\right)^{1/4}\\ C_e&=\left(\frac{A_V}{\pi_{1e}}\right)^{1/2}\left(\frac{A_F}{\pi_{2e}}\right)^{1/4}\left(\frac{A_s}{\pi_{3e}}\right)^{1/4}\end{aligned}\right\} \tag{3.74}$$

With some practice, the above system of equations can be written directly by inspection of the inverse dimension matrix.

To determine economical similarity criteria for Eq. (3.70) we shall use a relation known from linear algebra:

$$\|\alpha^T\|^{-1}=\|\alpha^{-1}\|^T$$

Then the tableau (3.73) can be converted to the following form:

$$\|\alpha^{-1}\|^T=\begin{Vmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{Vmatrix}$$

According to Eq. (3.68), similarity criteria for economical alternatives can be found by the equation

$$\begin{Vmatrix} \pi_{1e} \\ \pi_{2e} \\ \pi_{3e} \end{Vmatrix}=\begin{Vmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{Vmatrix}\cdot\begin{Vmatrix} 0 \\ 0 \\ -1 \end{Vmatrix}=\begin{Vmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{4} \end{Vmatrix} \tag{3.75}$$

and have the following values:

$$\pi_{1e}=1/2, \qquad \pi_{2e}=1/4, \qquad \pi_{3e}=1/4 \tag{3.75a}$$

It is seen from Eq. (3.75*a*) that the total costs of the transmission line in question are itemized as follows: for the economical design the cost proportional to the energy lost in the line, π_{3e}, is equal to the cost proportional to the conductor cross section, π_{2e}, and the sum of these two items is equal to the cost proportional to voltage, π_{1e}.

This itemized costing of an economical design can be shown graphically, as is done in the chart of Fig. 3.7. In this chart, the area of the circle, taken as unity, represents the total economical costs of the transmission line, and the areas of the sectors represent the individual cost items in the original model for the economical alternative. This break-down of the total costs will hold for any design data.

The similarity criteria of economic alternatives have been evaluated by use of tableau equation (3.75). However, these same similarity criteria can be found from the inverse dimension matrix α^{-1}. This is so because in the matrix α^{-1}, Eq. (3.75), only the last element β_n has a non-zero value. Therefore, the similarity criteria π_{ie} can be evaluated by multiplying the elements in the last column of the matrix $\|\alpha^T\|^{-1}$ by the element $\beta_n=-1$. So, the similarity criteria of economical alternatives are numerically equal to the elements in the last column of the matrix $\|\alpha^T\|^{-1}$, taken with the opposite sign. As already noted, however, $\|\alpha^T\|^{-1}=\|\alpha^{-1}\|^T$, and so the simi-

larity criteria can be determined from the last row of the matrix α^{-1}, if we change the signs of its elements.

Thus, we can use the inverse of the dimension matrix α not only to write the equations for economical alternatives, but also to evaluate the similarity criteria. As will be shown later, in this way we simultaneously solve both the primal and the dual problems of criterion programming.

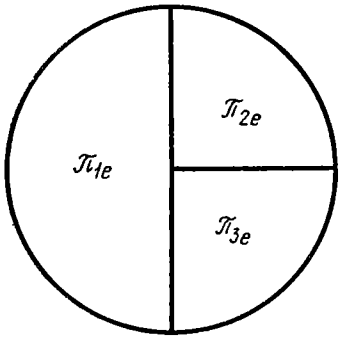


Fig. 3.7

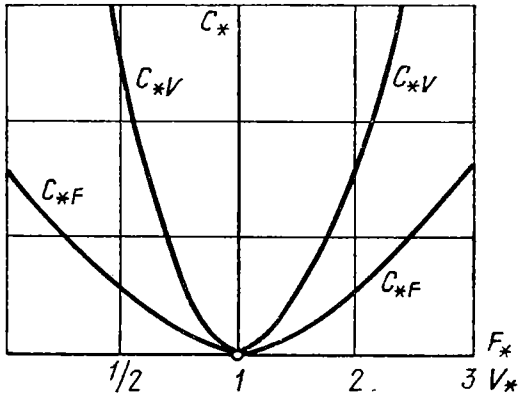


Fig. 3.8

Once we know the values of the similarity criteria, we can readily construct a criterion model of the line by the previously formulated rule: replace the constants of Eq. (3.70) by similarity criteria, and the variables and costs by their per-unit values:

$$C_* = V_*/2 + F_*/4 + 1/4 V_*^2 F_* \tag{3.76}$$

Now we can test this model for economical stability towards F_* and V_*. For this purpose we refer to Fig. 3.8 which is a plot of the equations:

$$C_{F*} = 1/2 + F_*/4 + F_*^{-1}/4$$
$$C_{V*} = 1/4 + V_*/2 + V_*^{-2}/4$$

It is seen from the plot that the criterion model of the transmission line, Eq. (3.76), is more sensitive towards departures in V, especially to its decrease. If, for some reason, we replace the economical cross section with one twice as great or one half as great, the relative increase in the variable cost item will only be 12.5%. If we double the voltage, the relative cost will rise by about 28%. If we halve the voltage, the relative increase in the cost will be as high as 50%.

Equation (3.76) is very general. Yet, the examples considered seem to suggest that special care in design work should be given to the economic analysis of the operating voltage to be chosen for distribution lines.

Example 3.3. Consider the electricity supply system of a metropolitan area built up with blocks of flats. The variables to be optimized will be the

service radius R of the metropolitan substation and the side L of the square serviced by one distribution point (DP). Since the load density per square kilometre of metropolitan area is constant, we shall thus find the economical loads on the elements considered.

If we choose the minimum of capital costs per kilowatt of power transmitted as a criterion of economical optimality, the design equation for the variable item of the capital costs will have the form

$$K=A_1R^{-2}+A_2RL^{-2}+A_3L \tag{3.77}$$

where the first term represents the capital costs of the H.V. bulk supply line and the district substation, the second term represents the capital cost of the M.V. supply line and the DP, and the third, the capital cost of the M.V. distribution networks.

On the basis of the design equation, we write a dimension matrix α and invert it:

$$\alpha=\begin{array}{c|ccc} & R & L & K \\ \hline \text{I} & -2 & 0 & -1 \\ \text{II} & 1 & -2 & -1 \\ \text{III} & 0 & 1 & -1 \end{array}, \qquad \alpha^{-1}=\begin{array}{c|ccc} & \text{I} & \text{II} & \text{III} \\ \hline R & -3/7 & 1/7 & 2/7 \\ L & -1/7 & -2/7 & 3/7 \\ K & -1/7 & -2/7 & -4/7 \end{array} \tag{3.78}$$

The economical similarity criteria can be evaluated from the last row of the inverse dimension matrix:

$$\pi_{1e}=1/7, \qquad \pi_{2e}=2/7, \qquad \pi_{3e}=4/7 \tag{3.79}$$

In words, the electricity supply system of the metropolitan area considered is economical, if 4/7 of the total variable specific capital cost is accounted for by the M.V. distribution networks, 2/7 by the M.V. bulk supply lines and the DP, and 1/7 by the district substation and the primary transmission line.

To determine an economical alternative from the inverse dimension matrix, we write the following equations:

$$\left.\begin{aligned} R_e&=2(A_1^3/4A_2A_3^2)^{1/7} \\ L_e&=2(A_1A_2^2/8A_3^3)^{1/7} \\ K_e&=\frac{7}{2}\left(\frac{1}{8}A_1A_2^2A_3^4\right)^{1/7} \end{aligned}\right\} \tag{3.80}$$

In per-unit form,

$$\left.\begin{aligned} R_*&=(A_{1*}^3/(A_{2*}A_{3*}^2)^{1/7} \\ L_*&=(A_{1*}A_{2*}^2/A_{3*}^3)^{1/7} \\ K_*&=(A_{1*}A_{2*}^2A_{3*}^4)^{1/7} \end{aligned}\right\} \tag{3.81}$$

Eqs. (3.80) make it possible to trace the effect of the constants A_1, A_2 and A_3 on the economical values of the variables, that is, to test the model for technicoeconomic sensitivity.

Referring to Eq. (3.77), the criterion equation may be written as:

$$K_* = R_*^{-2}/7 + 2R_* L_*^{-2}/7 + 4L_*/7 \tag{3.82}$$

which may be used to analyse the electricity supply system under consideration for stability as the variables depart from their economical values.

3.14. Dimension Matrix

In determining economical similarity criteria, variables and costs by Eqs. (3.59) and (3.68), we assumed that the dimension matrix was non-singular. It is only under this assumption that economical similarity criteria depend solely on the exponents and are independent of the numerical values of the constants A_i. Since this case is of important practical significance, it will be worth while analysing the dimension matrix in greater detail.

The dimension matrix α is singular, if the number of cost items m in the design equation, Eq. (3.49), is:

(1) less than the number of variables n plus one, that is, if $m < n+1$*;

(2) greater than the number of variables n plus one, that is, the number of variables being optimized plus costs: $m > n+1$;

(3) equal to the number of variables in the initial model, n plus one, but some rows or columns of the matrix α are linear combinations of the remaining rows or columns, that is, if the rank r of the matrix is lower than the number of cost items: $r < m$.

An example of the first case, where $m < n+1$, is the following equation:

$$C = A_1 x_1 x_2^{-2} + A_2 x_1^{-1} x_2^3, \qquad m=2, \qquad n=2 \tag{3.83}$$

which is reducible to canonical form if the statement of the problem allows the variable x_1 or x_2 to be absorbed in the constants A_1 and A_2:

$$C = B_1 x_2^{-2} + B_2 x_2^3 \tag{3.84}$$

$$C = D_1 x_1 + D_2 x_1^{-1} \tag{3.85}$$

The dimension matrix for Eqs. (3.83) through (3.85) has the form:

$$\alpha_1 = \begin{Vmatrix} 1 & -2 & -1 \\ -1 & 3 & -1 \end{Vmatrix} \quad \alpha_2 = \begin{Vmatrix} -2 & -1 \\ 3 & -1 \end{Vmatrix} \quad \alpha_3 = \begin{Vmatrix} 1 & -1 \\ -1 & -1 \end{Vmatrix}$$

Matrix α_1 is singular, while matrices α_2 and α_3 are non-singular.

An example of the second case, where $m > n+1$, is the equation

$$C = A_1 x^{\alpha_1} + A_2 x^{-\alpha_2} + A_3 x^{\alpha_3} \tag{3.86}$$

* In the notation adopted, n stands for the number of variables x being optimized. To include one more term, namely C, unity should be added.

for which the dimension matrix is

$$\alpha_4 = \begin{Vmatrix} \alpha_1 & -1 \\ -\alpha_2 & -1 \\ \alpha_3 & -1 \end{Vmatrix}$$

If the first or third terms of Eq. (3.86) cannot be ignored, the economical similarity criteria cannot be evaluated solely from the exponents alone, and knowledge of the constants A_1, A_2 and A_3 is necessary.

Now we shall evaluate the amount of information needed to determine the similarity criteria π_{ie} for $m > n+1$. Our analysis will be based on the general case characterized by Eq. (3.49).

As has already been shown, the similarity criteria of economical alternatives* satisfy a matrix equation of the form

$$\alpha^T \pi = \beta$$

which can be written at length as

$$\left.\begin{aligned} \alpha_{11}\pi_1 + \ldots + \alpha_{m1}\pi_m &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ \alpha_{1n}\pi_1 + \ldots + \alpha_{mn}\pi_m &= 0 \\ -\pi_1 - \ldots - \pi_m &= -1 \end{aligned}\right\} \tag{3.87}$$

We leave $(n+1)$ terms on the left-hand side of the above equation and call them the primary terms. We transpose the remaining terms $m-(n+1)$ to the right-hand side and call them the auxiliary terms. In so doing, we use the following designations:

$$\alpha_p^T = \begin{Vmatrix} \alpha_{11} \cdots \alpha_{n+1,1} \\ \cdots\cdots\cdots \\ \alpha_{1n} \cdots \alpha_{n+1,n} \\ -1 \ldots -1 \end{Vmatrix}, \quad \alpha_a^T = \begin{Vmatrix} \alpha_{n+2,1} \cdots \alpha_{m1} \\ \alpha_{n+2,n} \cdots \alpha_{mn} \\ -1 \quad \ldots -1 \end{Vmatrix}$$

$$\pi_p = \begin{Vmatrix} \pi_1 \\ \vdots \\ \pi_{n+1} \end{Vmatrix}, \quad \pi_a = \begin{Vmatrix} \pi_{n+2} \\ \vdots \\ \pi_m \end{Vmatrix}, \quad \beta = \begin{Vmatrix} 0 \\ \vdots \\ 0 \\ -1 \end{Vmatrix}$$

Then, Eqs. (3.87) may be re-written as

$$\alpha_p^T \pi_p = \beta - \alpha_a^T \pi_a \tag{3.88}$$

Assume that the matrix α_p^T is non-singular. Otherwise, we would have to interchange the terms of Eqs. (3.87) so as to make the matrix α_p^T non-

* In this example, the subscript 'e' of the economical similarity criteria has been dropped for simplicity.

singular. Premultiplying (3.88) by the inverse matrix $\|\alpha_p^T\|^{-1}$ gives the dependence of the primary similarity criteria π_p on the surplus (auxiliary) similarity criteria π_a:

$$\pi_p = \|\alpha_p\|^T \beta - \|\alpha_p^T\|^{-1} \alpha_a^T \pi_a \tag{3.89}$$

To evaluate the auxiliary similarity numbers, we shall use Eqs. (3.52) which hold for the general case

$$\pi_a = \frac{A_t}{C_e} \prod_{j=1}^{n} x_{je}^{\alpha_{tj}}; \qquad t = \begin{vmatrix} m \\ n+2 \end{vmatrix}$$

Now we divide the equations by A_t and write them in contracted matrix form

$$\mathbf{K}_a = \alpha_a \mathbf{Y}_e \tag{3.90}$$

where

$$\mathbf{K}_a = \left\| \begin{matrix} \ln \dfrac{\pi_{n+2}}{A_{n+2}} \\ \vdots \\ \ln \dfrac{\pi_m}{A_m} \end{matrix} \right\|, \qquad \mathbf{Y}_e = \left\| \begin{matrix} \ln x_{1e} \\ \vdots \\ \ln x_{ne} \\ \ln C_e \end{matrix} \right\|$$

Substituting the economic value of $\mathbf{Y}_e$ from Eq. (3.59) in Eq. (3.90) gives

$$\mathbf{K}_a = \alpha_a - \alpha_p^{-1} \mathbf{K}_e = \lambda \mathbf{K}_e$$

After we take the anti-logarithms, we can pass to the original variables:

$$\frac{\pi_t}{A_t} = \left[\left(\frac{\pi_{1,0}}{A_1} \right)^{\lambda_{t,1}} \cdots \left(\frac{\pi_{n+1,0}}{A_{n+1}} \right)^{\lambda_{t,n+1}} \right]^{1/\Delta_0}, \qquad t = \begin{vmatrix} m \\ n+2 \end{vmatrix}$$

whence

$$\pi_t = A_t \prod_{i=1}^{n+1} \left(\frac{\pi_{i0}}{A_i} \right)^{\lambda_{ti}/\Delta_0}, \qquad t = \begin{vmatrix} m \\ n+2 \end{vmatrix} \tag{3.91}$$

Thus, in the second case, where $m > n+1$, the economical similarity criteria depend not only on the exponents a_{ij} of the variables being optimized, but also on the constants A_t. They are evaluated by simultaneously solving $(n+1)$ linear equations (3.89) and $m-(n+1)$ nonlinear equations (3.91).

The third case, where $m = n+1$ and $r < m$, can readily be reduced to the first or second case and is therefore of no particular interest.

To sum up, analysis of the dimension matrix shows that although it is possible to determine the economical similarity criteria of non-canonical equations for $m > n+1$, the analytical relations thus obtained are not so graphic as they are when the equations are canonical. Besides, it is necessary to prove that the economical alternative is existent and unique.

In problems involving non-canonical equations, other methods may be used for determining economical alternatives. However, none of them appears most effective or able to take proper account of the specific features of a particular discounted cost equation.

3.15. Parametric Similarity

In selecting an economical cross section for the conductors of a power transmission line by the equation

$$C=A_F F+A_s/F \tag{3.92}$$

it was shown that Kelvin's similarity criterion is equal to unity:

$$k_{Fe}=\pi_{se}/\pi_{Fe}=1$$

If the economical cross sections of line conductors be selected simultaneously with an economical voltage for the line by the equation

$$C=A_V V+A_F F+A_s/(V^2F) \tag{3.93}$$

the economical similarity criteria will have the following values

$$\pi_{Ve}=1/2, \qquad \pi_{Fe}=1/4, \qquad \pi_{se}=1/4$$

At the same time, Kelvin's criterion which relates the cost items associated with the variable F will remain equal to unity, although the primary similarity criteria π_{se} and π_{Fe} have changed in magnitude.

The fact that the secondary similarity criterion, k_{Fe}, retains its value is due to what is called *parametric similarity*. The latter arises because the variable F occurs in only two terms of Eqs. (3.92) and (3.93) and its exponent is the same in both cases.

It can be shown that parametric similarity is preserved in the general case, too. Suppose that the variable to be optimized, x_k, appears in only two terms of the equation

$$C=\sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_j^{\alpha_{ij}}$$

with the index $i=s$ and $i=t$. Equating to zero the first derivative of C with respect to x_k yields

$$\frac{\partial C}{\partial x_k}=\frac{\alpha_{sk}A_s}{x_{ke}}\prod_{j=1}^{n} x_{je}^{\alpha_{sj}}+\frac{\alpha_{tk}A_t}{x_{ke}}\prod_{j=1}^{n} x_{je}^{\alpha_{tj}}=0 \tag{3.94}$$

Since

$$A_i \prod_{j=1}^{n} x_{je}^{\alpha_{ij}}=\pi_{ie}C_e$$

the condition (3.65) can be re-written as

$$\alpha_{sk}\pi_{se}+\alpha_{tk}\pi_{te}=0$$

whence

$$k_{Fe}=\pi_{te}/\pi_{se}=-\alpha_{sk}/\alpha_{tk}=\text{idem}$$

The above criterion equation is a proof that parametric similarity in x_k is preserved, regardless of whether the design equation is in canonical form or

not. What matters in the case of parametric similarity is the canonical form with respect to the variable concerned.

Parametric similarity may be utilized in forecasting. We shall illustrate this possibility with an example.

Example 3.4. Determine the economical quantity of conductor material for distribution networks designed to transmit 2000 TWh of electric energy.

The forecast for the period in question has shown that no major breakthroughs in the methods of electric power transmission are to be expected. Then, during the period under observation, as at present, the variable items of the line cost will be determined by Eq. (3.92) written as

$$C=A_F Q+\Delta A b \tag{3.95}$$

where Q=quantity of conductor material
ΔA=transmission loss
b=cost of kWh lost

In view of the fact that Kelvin's criterion remains constant, we may rewrite Eq. (3.95) as follows:

$$\Delta A b/A_F Q=k_{Fe}=1$$

Hence, the weight of conductor material required will be

$$G=dQ=d\Delta A b/A_F$$

where d is the density of the conductor material.

Notably, for $d=2.7$ t/m^3, $\Delta A=10\%$, $b=0.5\times10^{-2}$ roubles/kWh, $A_F=5$ roubles/km mm^2, we get:

$$G=2.7\times0.1\times2\times10^{12}\times0.5\times10^{-5}=0.54\times10^6 \text{ tonnes}$$

Or, in words, assuming that the design cost data remain constant, the transmission line needed to handle 2000 TWh of electricity will use 540,000 tonnes of aluminium as conductor material. As is seen, this calculation, approximate as it may be, gives a preview of how much aluminium will be needed without resort to specific network calculations.

3.16. First and Second Order Invariance

Parametric similarity is closely related to a special form of the invariance of economical similarity criteria, called *second-order invariance*. In brief, this concept may be outlined as follows.

If the design equation is in canonical form, the economical similarity criteria will only depend on the exponents α_{ij} and be independent of the constants A_i. This property of similarity criteria is called *first-order invariance*. This form of invariance is at the basis of economic stability studies of canonical models by use of the criterion equation. In view of similarity

between economical alternatives, we may generalize the criterion equation to a multitude of installations having different sets of constants A_i.

In practical technico-economic analysis, the designer often encounters second-order invariance, where the economical similarity criteria remain constant with changes in both the constants A_i and the exponents α_{ij}.

As an example, we shall analyse the following equations:

$$\left.\begin{aligned} C &= A_1 x_1^2 x_2^1 + A_2 x_1^{-1} x_2^{-1} + A_3 x_1^1 x_2^2 \\ Z &= B_1 y_1^3 y_2^{-1} + B_2 y_1^{-2} + B_3 y_1^3 y_2^1 \end{aligned}\right\} \tag{3.96}$$

The dimension matrices and their inverses for the above equations have the form

$$\alpha_C = \begin{Vmatrix} 2 & 1 & -1 \\ -1 & -1 & -1 \\ 1 & 2 & -1 \end{Vmatrix}, \qquad \alpha_Z = \begin{Vmatrix} 3 & -1 & -1 \\ -2 & 0 & -1 \\ 3 & 1 & -1 \end{Vmatrix}$$

$$\alpha_C^{-1} = \begin{Vmatrix} 3/5 & -1/5 & -2/5 \\ -2/5 & -1/5 & 3/5 \\ -1/5 & -3/5 & -1/5 \end{Vmatrix}, \qquad \alpha_Z^{-1} = \begin{Vmatrix} 1/10 & -1/5 & 1/10 \\ -1/2 & 0 & 1/2 \\ -1/5 & -3/5 & -1/5 \end{Vmatrix}$$

From the numerical values of the inverse matrices it follows that the economical similarity criteria will be the same in both equations for any numerical values of the constants A_i and B_i and the exponents α_{ij}:

$$\pi_{1e} = 1/5, \qquad \pi_{2e} = 3/5, \qquad \pi_{3e} = 1/5 \tag{3.97}$$

By the same token, the respective sides of Eqs. (3.96) are likewise similar

$$\left.\begin{aligned} \pi_{1e} &= A_1 x_{1e}^2 x_{2e}^1 / C_e = B_1 y_{1e}^3 y_{2e}^{-1} / Z_e = 1/5 \\ \pi_{2e} &= A_2 x_{1e}^{-1} x_{2e}^{-1} / C_e = B_2 y_{1e}^{-2} / Z_e = 3/5 \\ \pi_{3e} &= A_3 x_{1e}^1 x_{2e}^2 / C_e = B_3 y_{1e}^3 y_{2e}^1 / Z_e = 1/5 \end{aligned}\right\} \tag{3.98}$$

Using the inverse matrix, the following equations for economical costs may be written:

$$C_e = 5(A_1 A_2^3 A_3)^{1/5}$$
$$Z_e = 5(B_1 B_2^3 B_3)^{1/5}$$

If $A_1 = B_1$, $A_2 = B_2$ and $A_3 = B_3$, then $C_e = Z_e$, and Eqs. (3.98) may be simplified as follows:

$$x_{1e}^2 x_{2e} = y_{1e}^3 y_{2e}^{-1}$$
$$x_{1e}^{-1} x_{2e}^{-1} = y_{1e}^{-2}$$
$$x_{1e} x_{2e}^2 = y_{1e}^3 y_{2e}$$

Solving the above equations simultaneously yields the following economical values of the variables:

$$x_{1e} = y_{1e}/y_{2e}, \qquad x_{2e} = y_{2e}, \qquad y_{1e} = x_{1e} x_{2e}$$

Thus, where equality (3.98) holds, simple relationships can be established between the economical values of the objective variables having the same constants.

3.17. Discrete Constraints

In the general case, the specific values of the objective variables (those being optimized) are found subject to the constraints imposed on these variables.

Using a three-term canonical equation similar to Eq. (3.70) as an example, we shall examine how discrete constraints can affect criterion analysis.

To begin with, we divide the feasible area of each of the two objective variables into three ranges in accordance with Table 3.1 (Fig. 3.9).

Table 3.1

Ranges of Objective Variables

Variable x_1	Variable x_2		
	$x_2 < x_{2\,\min}$	$x_{2\,\min} \leqslant x_2 \leqslant x_{2\,\max}$	$x_2 > x_{2\,\max}$
$x_1 < x_{1\,\min}$	11	12	13
$x_{1\,\min} \leqslant x_1 \leqslant x_{1\,\max}$	21	22	23
$x_1 > x_{1\,\max}$	31	32	33

Assume that the economical values of the variables lie in range *22* (point *e*). This value is associated with minimum costs, C_e. If the variables x_1 and x_2 have values other than economical, the respective cost values will rise. The aggregate of all likely cost values in a three-dimensional space will be depicted by a surface resembling an umbrella turned upside down.

If x_1 and x_2 vary within the feasible range continually, the optimal design is an economical one, $x_{10}=x_{1e}$, $x_{20}=x_{2e}$, and $C_0=C_e$.

If x_1 and x_2 can take each only two discrete values, $x_{1\,\min}$ and $x_{1\,\max}$ and, respectively, $x_{2\,\min}$ and $x_{2\,\max}$, the optimal alternative will be that for which the costs (C_A, C_C, C_E or C_G) at the respective point (A, C, E or G) are a minimum.

The procedure will be the same if the range of a variable contains more than two values. Where a more complex problem is encountered, resort may be made to some concepts of integer programming. This brings us to the need for integer criterion programming which would utilize the merits of both criterion analysis and integer programming.

Consider a case where the economical alternative occurs in the remaining eight ranges.

If the economical values of the objective variables occur in ranges *11* and *13* or *31* and *33*, the costs associated with the optimal alternatives will be C_G, C_A, C_C or C_E.

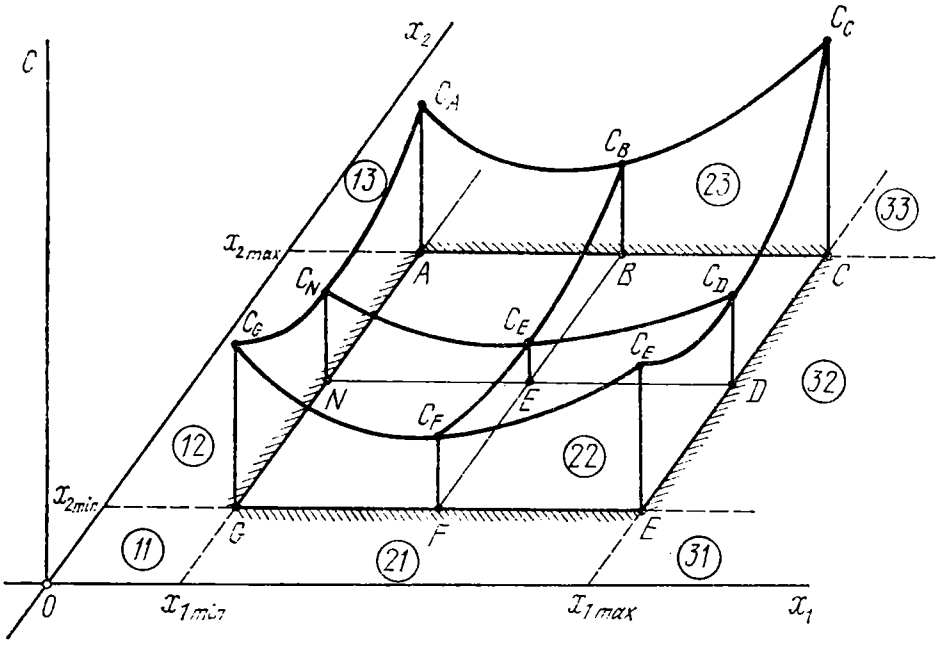


Fig. 3.9

Assume also that the economical values of the objective variables occur in range *12* (Fig. 3.9). We pass horizontal planes through the discounted cost surface and construct projections of the lines of intersections on the $x_1 - o - x_2$ plane (Fig. 3.10).

It is always possible to choose a plane such that the projection of the line of intersection touches (at point *N*) the straight line *AG* which bounds the feasible area of the variables. The coordinates of point *N* will then

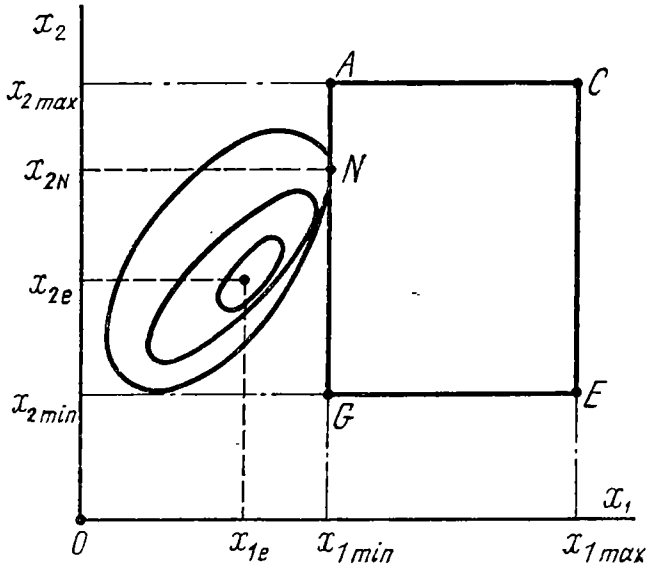


Fig. 3.10

define the optimal alternative, because they satisfy the original constraints. In improving the values of x_{10} and x_{20}, it should be recalled that the variables are constrained to take a set of discrete values.

When the economical values of the objective variables occur in range *21*, *23* or *32*, the optimal alternative is defined in an analogous way.

3.18. Ellipsoids of Equal Precision

As already noted, criterion equations make it possible to determine the relative increase in costs due to a relative departure of the variables from their economical values. In design work, however, it is often necessary to handle the reverse problem, that is, to determine the allowable limits of change in the objective variables for a specified (prescribed) relative increase in costs.

In such cases, a new concept has to be introduced for canonical equations — an ellipsoid of equal precision. In brief, this concept can be explained as follows.

For $m=n+1$, the canonical criterion equations are written as

$$C_*=\sum_{i=1}^{n+1}\pi_i\prod_{j=1}^{n}x_{j*}^{\alpha_{ij}} \tag{3.99}$$

Geometrically, C_* is a surface in the $(n+1)$-ary space. Let us explore this surface in the neighbourhood of the minimum point. In doing so, we shall use the Taylor theorem in order to approximate the C_* surface with what is called a tangent paraboloid which reproduces the $C_*(x_{j*})$ surface locally in much the same way as a line tangent to a plane reproduces the surface in the neighbourhood of the point of tangency.

By the Taylor theorem, a continuous and doubly differentiable n-ary scalar function, $f(x)$, defined on the interval $[a,\ a+\varepsilon]$, may be expressed as a finite series in powers of ε:

$$f(a+\varepsilon)=f(a)+\frac{\varepsilon}{1!}f'(a)+\frac{\varepsilon^2}{2!}f''(a) \tag{3.100}$$

where $\varepsilon=x-a=$increment in the variable

$f'(a), f''(a)=$first and second derivatives of $f(x)$ with respect to x

To utilize the above expansion for an approximate representation of criterion equation (3.99), we evaluate the first and second derivatives:

$$\frac{\partial C_*}{\partial x_{k*}}=\sum_{i=1}^{n+1}\frac{\alpha_{ik}\pi_i}{x_{k*}}\prod_{j=1}^{n}x_{j*}^{\alpha_{ij}}, \qquad k=\overline{|}^{n}_{1}$$

$$\frac{\partial^2 C_*}{\partial x_{k*}\,\partial x_{t*}}=\sum_{i=1}^{n+1}\frac{\alpha_{ik}\alpha_{it}\pi_i}{x_{k*}x_{t*}}\prod_{j=1}^{n}x_{j*}^{\alpha_{ij}}, \qquad k,\ t=\overline{|}^{n}_{1}$$

Designating

$$\beta_{tk}=\beta_{kt}=\sum_{i=1}^{n+1}\alpha_{ik}\alpha_{it}\pi_i, \qquad k,\ t=\overline{|}^{n}_{1} \tag{3.101}$$

and recalling that at the point $x_{j*}=1$ the first derivatives vanish because

$$\sum_{i=1}^{n+1} \alpha_{ik}\pi_i=0, \qquad k=\left|\begin{matrix} n \\ 1 \end{matrix}\right.$$

the criterion equation, Eq. (3.99), may be approximated by the equation

$$C_* \approx 1+\frac{1}{2}\sum_{k,\,l=1}^{n} \beta_{kl}(x_{k*}-1)(x_{l*}-1)$$

which describes an n-ary paraboloid (Fig. 3.11). We replace the relative terms in the above equation with their relative increments:

$$\varepsilon_*=x_*-1$$
$$\varepsilon_C=C_*-1$$

Then

$$2\varepsilon_C=\sum_{k,\,l=1}^{n} \beta_{kl}\varepsilon_{k*}\varepsilon_{l*} \tag{3.102}$$

This change of variable implies the transfer of coordinate axes. By letting the cost plane cut the paraboloid, we obtain an n-ary surface, called an *ellipsoid.* If the relative increments in the variables lie within the ellipsoid,

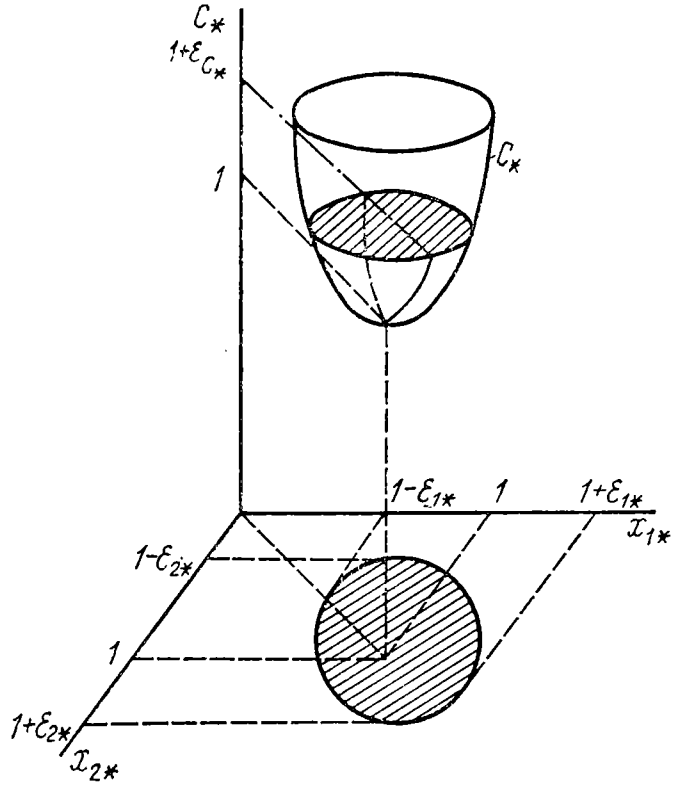


Fig. 3.11

it may be expected that the costs will not exceed ε_C. In other words, the ellipsoid defines the area of relative departures in the objective variables from their optimal values, consistent with the specified (prescribed) limit of increase in the costs. This is why we call it an *ellipsoid of equal precision.*

Equation (3.102) is a quadratic one in the variables ε_*. It is known from linear algebra that by use of orthogonal transformation a quadratic equation can be presented as the sum of the squares of the variables, with the constants of the squares being the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the matrix of coefficients β_{kl}. Geometrically, this implies that the coordinate axes can be rotated until they are coincident with the axes of the respective ellipsoid.

Economically, this rotation makes it possible to define the limits of departures in the objective variables for a specified (prescribed) limit of error in costs. This is why it is of practical interest to determine the axes of the ellipsoid. We shall demonstrate how this can be done in the general case.

The number λ can be evaluated by solving the characteristic equation of degree n:

$$\begin{vmatrix} \beta_{11}-\lambda & \beta_{12} & \ldots & \beta_{1n} \\ \beta_{21} & \beta_{22}-\lambda & \ldots & \beta_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \beta_{n1} & \beta_{n2} & \ldots & \beta_{nn}-\lambda \end{vmatrix} = 0 \tag{3.103}$$

It can be shown that all n roots of the characteristic equation, Eq. (3.103), are real and positive, $\lambda_i > 0$. We arrange these numbers in ascending order:

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n \tag{3.104}$$

Once the eigenvalues λ_i are known, we may write the quadratic equation, Eq. (3.102), as

$$\lambda_1 \xi_1^2 + \lambda_2 \xi_2^2 + \ldots + \lambda_n \xi_n^2 = 2\varepsilon_C$$

or

$$\frac{\xi_1^2}{a_1^2} + \frac{\xi_2^2}{a_2^2} + \ldots + \frac{\xi_n^2}{a_n^2} = 1 \tag{3.105}$$

where

$$a_i = (2\varepsilon_C/\lambda_i)^{1/2}, \qquad i = \overline{1,n} \tag{3.106}$$

We can go from the ξ coordinates to the ε_* coordinates by the following matrix transformation:

$$\varepsilon_{i*} = \sum_{j=1}^{n} q_{ij}\xi_i, \qquad i = \overline{1,n} \tag{3.107}$$

where the coefficients q form an orthogonal transformation matrix Q.

Equation (3.105) is canonical in defining the axes of the ellipsoid of equal precision, because the axes are equal to the corresponding numbers

a_i found by Eqs. (3.106). According to condition (3.104), the direction of maximum economical stability is associated with the minimum eigenvalue λ_i.

It is to be noted that for canonical functions the direction of the axes is independent of the constants and is solely decided by the exponents α_{ij}.

As an example of constructing an ellipsoid of equal precision, we shall consider a canonical equation of the form

$$C=A_1F+A_2R^2F^{-1}+A_3R^{-1}$$

which describes the cost of an L.V. network as a function of conductor cross section F and service radius R. The economical values of the similarity criteria for this equation are:

$$\pi_{1e}=1/4, \qquad \pi_{2e}=1/4, \qquad \pi_{3e}=1/2$$

The criterion equation may then be written as

$$C_*=F_*/4+R_*^2F_*^{-1}/4+R_*^{-1}/2$$

The coefficients β_{kl} can be evaluated from Eq. (3.101)

$$\beta_{11}=1^2\times\frac{1}{4}+(-1)^2\frac{1}{4}+0^2\times\frac{1}{2}=\frac{1}{2}$$

$$\beta_{12}=\beta_{21}=1\times0\times\frac{1}{4}+(-1)(2)\frac{1}{4}+0(-1)\frac{1}{2}=-\frac{1}{2}$$

$$\beta_{22}=0^2\times\frac{1}{4}\times2^2\times\frac{1}{4}+(-1)^2\frac{1}{2}=\frac{3}{2}$$

Thus, the characteristic equation may be written in the form of a determinant, Eq. (3.103)

$$\begin{vmatrix} \frac{1}{2}-\lambda & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2}-\lambda \end{vmatrix}=0$$

or

$$\lambda^2-2\lambda+1/2=0$$

whence

$$\lambda_1=1-1/\sqrt{2}$$

$$\lambda_2=1+1/\sqrt{2}$$

Therefore, the semi-axes of the ellipse described by Eq. (3.106) may be defined as

$$a_1=\left(\frac{2\varepsilon}{1-1/\sqrt{2}}\right)^{1/2}$$

$$a_2=\left(\frac{2\varepsilon}{1+1/\sqrt{2}}\right)^{1/2}$$

Assuming that the limit of error is 5%, then $\varepsilon=0.05$, and so

$$a_1=0.586, \qquad a_2=0.242$$

Now the ellipse can be described by an equation of the form of (3.105):

$$\xi_1^2/0.586^2+\xi_2^2/0.242^2=1 \tag{3.108}$$

To go back to the original coordinates F_* and R_*, we should first evaluate the orthogonal transformation matrix

$$\mathbf{Q}=\begin{Vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{Vmatrix} \tag{3.109}$$

The numbers of this matrix should satisfy the system of homogeneous equations

$$\left.\begin{aligned}(\beta_{11}-\lambda)p+\beta_{12}q=0\\ \beta_{12}p+(\beta_{22}-\lambda)q=0\end{aligned}\right\} \tag{3.110}$$

once, at $\lambda=\lambda_1$, and for a second time at $\lambda=\lambda_2$. In the former case,

$$\left(-\frac{1}{2}+\frac{1}{\sqrt{2}}\right)p-\frac{1}{2}q=0$$

$$-\frac{1}{2}p+\left(\frac{1}{2}+\frac{1}{\sqrt{2}}q\right)=0$$

The second equation is a corollary of the first, and so, putting $p=1$, we find

$$q=\sqrt{2}-1$$

Next we normalize the eigenvector $l_1=\{1,\ \sqrt{2}-1\}$ of the matrix

$$\mathbf{B}=\begin{Vmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{Vmatrix}$$

corresponding to the eigenvalues λ_1, as follows:

$$l_1^0=\frac{l_1}{|l_1|}=\left\{\frac{1}{\sqrt{4-2\sqrt{2}}}, \qquad \frac{\sqrt{2}-1}{\sqrt{4-2\sqrt{2}}}\right\}$$

Solving the equations in a similar way for $\lambda=\lambda_2$ yields a partial solution

$$p=-1$$

$$q=\sqrt{2}+1$$

Therefore, the eigenvector $l_2=\{-1, \sqrt{2}+1\}$ corresponding to the eigenvalue λ_2, after normalizing will be written as

$$l_2^0=l_2/|l_2|=\{-1/\sqrt{4+2\sqrt{2}}, \quad \sqrt{2}+1/\sqrt{4+2\sqrt{2}}\}$$

Using the orthonormalized vectors l_1^0 and l_2^0, the orthogonal transformation matrix (3-109) may be written as

$$\mathbf{Q}=\left\|\begin{matrix}\dfrac{1}{\sqrt{4-2\sqrt{2}}} & -\dfrac{1}{\sqrt{4-2\sqrt{2}}} \\ \dfrac{\sqrt{2}-1}{\sqrt{4-2\sqrt{2}}} & \dfrac{\sqrt{2}+1}{\sqrt{4-2\sqrt{2}}}\end{matrix}\right\|$$

Using this matrix, we can go from coordinates ξ_1, ξ_2 to coordinates ε_F and ε_R:

$$\left\|\begin{matrix}\varepsilon_F \\ \varepsilon_R\end{matrix}\right\|=\left\|\begin{matrix}p_1 & p_2 \\ q_1 & q_2\end{matrix}\right\|\cdot\left\|\begin{matrix}\xi_1 \\ \xi_2\end{matrix}\right\| \tag{3.111}$$

and, as a consequence, to the original coordinates

$$F_*=1+\varepsilon_F$$
$$R_*=1+\varepsilon_R$$

This process is illustrated in Fig. 3.12. The best way to go from the coordinates ξ_1, ξ_2 to the coordinates ε_F, ε_R is by rotating the axes through an angle γ which can be found from the relation

$$\tan\gamma=q_1/p_1=(\sqrt{2}-1)/1$$
$$\gamma=22.5^\circ$$

Then, instead of Eq. (3.111), we may use the equations

$$\varepsilon_F=\xi_1\cos\gamma-\xi_2\sin\gamma$$
$$\varepsilon_R=\xi_1\sin\gamma+\xi_2\cos\gamma$$

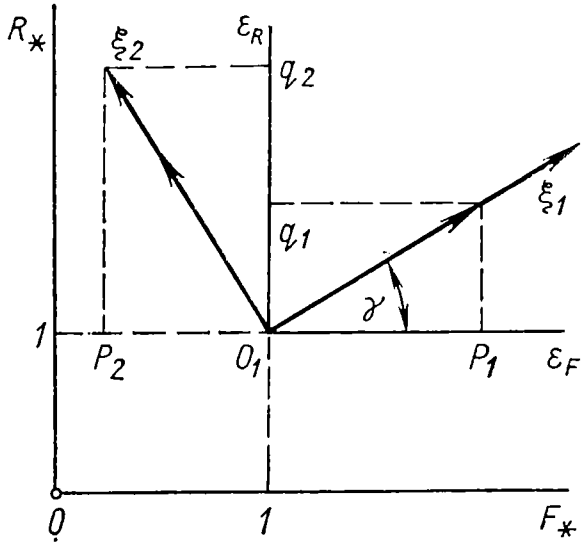


Fig. 3.12

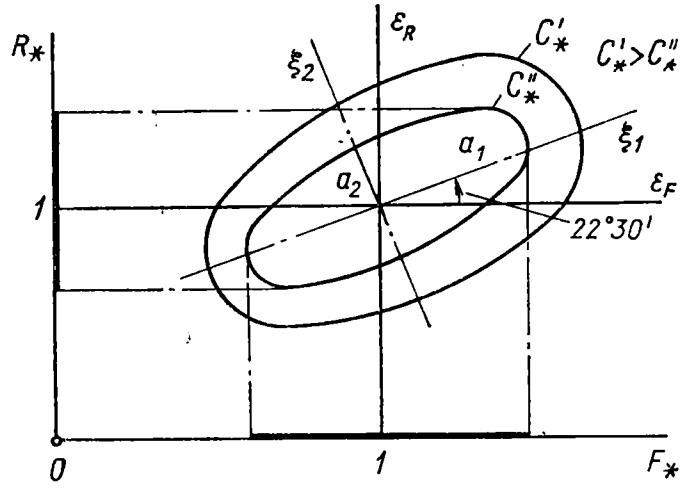


Fig. 3.13

For canonical functions, the orientation of the ellipse of equal precision (that is, the angle γ) is solely decided by the exponent α_{ij}, and the size of the axes by the limit of error in the costs, C_*. Fig. 3.13 shows two ellipsoids, one for C'_* and the other for C''_*, such that $C'_*>C''_*$. It is seen that an increase in C_* leads to an increase in the ellipse of equal precision and, as a consequence, in the limits for departures of the variables from their economical values.

It has thus been shown that the ellipsoid of equal precision technique makes it possible to determine how much the objective variables may depart from their economical values without causing the discounted costs to exceed their prescribed limit.

3.19. Functional Constraints

The manner in which the functional constraints imposed on the objective variables (those being optimized) can be taken into account will be illustrated by a simple example where the original model has the form

$$C=A_1F+A_2F^{-1}R^2+A_3R \tag{3.112}$$

subject to

$$F^{-1}R^2\leqslant B \tag{3.113}$$

Equations (3.112) and (3.113) state that the cross-sectional area F and the service radius R should be selected so that the voltage drop on the line will not exceed a prescribed limit.

Using the constraint, Eq. (3.113), in the equality form, we can drop one of the objective variables, F or R. In the former case, substitution of (3.113) in (3.112) gives

$$C_R=A_1B^{-1}R^2+A_2B+A_3R^{-1} \tag{3.114}$$

and in the latter case,

$$C_F=A_1F+A_2B+A_3B^{-1/2}F^{-1/2} \tag{3.115}$$

The first equation can be used to evaluate the economical service radius

$$R_e = [A_3 B/(2A_1)]^{1/3} \tag{3.116}$$

and the second to evaluate the economical cross-sectional area

$$F_e = [A_3^2/(4A_1^2 B)]^{1/3} \tag{3.117}$$

Substituting the economical values of the variables, (3.116) and (3.117), in Eqs. (3.114) and (3.115) respectively yields the minimum cost

$$C_e = 3(A_3^2 A_1/4B)^{1/3} + AB_2 \tag{3.118}$$

However, Eq. (3.118) may be treated as one in which changes occur in B, that is, in the limit of the constraints imposed on the variables F and R. There is a competition with regard to B, and so there should be an economical value, B_e, which can be found from Eq. (3.118):

$$B_e = [A_3^2 A_1/(4A_2^3)]^{1/4}$$

Substitution of B_e in Eq. (3.118) gives the minimum cost

$$C_e = (64 A_1 A_2 A_3^2)^{1/4}$$

corresponding to the minimum cost in the model described by Eq. (3.112), for the economical values of the two variables, F_e and R_e.

Now we evaluate the economical similarity criteria for Eq. (3.118):

$$\pi_{1e} = 3/4, \qquad \pi_{2e} = 1/4$$

Then the criterion equation may be written as

$$C_{*e} = 3B_*^{-1/3}/4 + B_*/4$$

a plot of which appears in Fig. 3.14. If we express the economical values

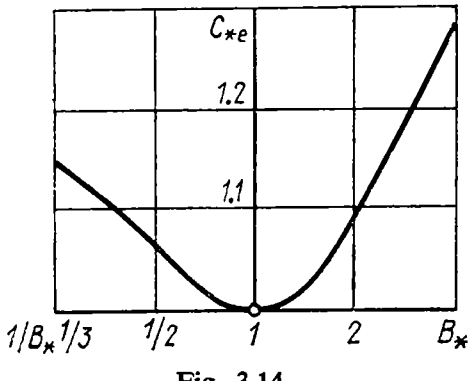


Fig. 3.14

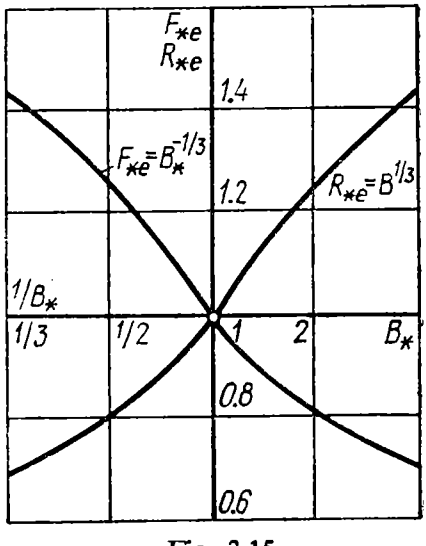


Fig. 3.15

of the variables, F_e and R_e, in per-unit form, too, then, by Eqs. (3.116) and (3.117), we get

$$R_{*e}=B_*^{1/3}$$

$$F_{*e}=B_*^{-1/3}$$

A plot of these relations is shown in Fig. 3.15.

In the above example, we have been able to carry out criterion analysis subject to functional constraints and have obtained fairly simple expressions. However, in the general case, functional constraints still pose a very complex and yet unsolved problem. Some special cases of the problem will be taken up in Sec. 3.28.

3.20. Equation with Negative Generalized Constants

The discounted cost equations examined so far have one feature in common — their generalized constants are positive, $A_i>0$. However, the engineer has often to deal with equations in which some constants are negative. Such equations are obtained, for example, in determining the economical degree of reactive power compensation in distribution networks.

Suppose there is a power transmission line operating at voltage V and having resistance R, which transmits electric energy to a consumer at point *2* (Fig. 3.16); the load receives both active and reactive power, varying in time, Q_{Lt}. If no reactive-power compensator is installed, $Q_c=0$,

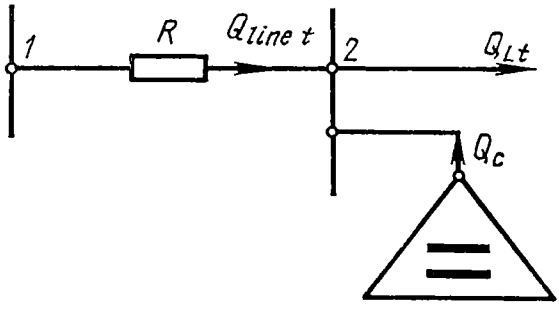


Fig. 3.16

the cost of reactive-power generation and transmission over time T will in a first approximation be

$$C_1=\int_0^T Q_{Lt}^2(Rb/V^2)\,dt+c_1\int_0^T Q_{Lt}\,dt$$

where b=cost per kWh at point *1*
c_1=cost per kVArh at point *1*

If we install an unregulated reactive-power compensator, RPC, of capacity Q_c, the cost of generation and transmission will be

$$C_2=\int_0^T Q_{line\,t}(Rb/V^2)\,dt+c_1\int_0^T Q_{line\,t}\,dt+c_cQ_cT$$

where c_c is the per-unit cost of reactive power supplied by the reactive-power compensator.

Provision of a reactive power compensator will be economically warranted if

$$E=C_1-C_2>0$$

Substituting in the last expression the values of C_1 and C_2 and also noting that

$$Q_c+Q_{line\,t}=Q_{Lt}$$

after some manipulations we get

$$E=\left[\frac{2Q_LRbT}{V^2}-(c_c-c_1)T\right]Q_c-(RbT/V^2)Q_c^2$$

where

$$Q_L=\frac{1}{T}\int_0^T Q_{LT}\,dt$$

Introducing generalized constants

$$A_1=2Q_LRbT/V^2-(c_c-c_1)T$$
$$A_2=RbT/V^2$$

the equation defining the optimum capacity of a reactive power compensator can be written as

$$E=A_1Q_c-A_2Q_c^2 \tag{3.119}$$

In Eq. (3.119), the constants A_1 and A_2 take opposite signs. This is why there is a competition between the economy terms. The similarity criteria for an economical alternative can be evaluated from the orthogonality and normality conditions:

$$\pi_{1e}+2\pi_{2e}=0$$
$$-\pi_{1e}-\pi_{2e}=-1$$

The dimension matrix has the form

$$\alpha=\begin{Vmatrix}1 & -1\\ 2 & -1\end{Vmatrix}$$

Its inverse is

$$\alpha^{-1}=\begin{Vmatrix}-1 & 1\\ 2 & 1\end{Vmatrix}$$

whence the similarity criteria of the economical alternative are

$$\pi_{1e}=2, \qquad \pi_{2e}=-1$$

As is seen, although the similarity criteria satisfy the orthogonality and normality conditions, one of them is greater than unit and the other is negative.

Using the inverse dimension matrix, we may write equations to determine the economical alternative as

$$\left.\begin{aligned} Q_{c.e}&=\left(\frac{-2}{A_1}\right)^{-1}\left(\frac{1}{A_2}\right)^{1}=Q_L-\frac{(c_c-c_1)}{2b}\cdot\frac{V^2}{R}\\ E_{\max}&=\left(\frac{-2}{A_1}\right)^{-2}\left(\frac{1}{A_2}\right)^{1}=\frac{Q_{c.e}^2}{V^2}RbT \end{aligned}\right\} \tag{3.120}$$

The term

$$\frac{(c_c-c_1)}{2b}(V^2/R)=Q_{line\,e}$$

is the economical value of reactive power that should be transmitted to the receiving end of the power transmission line. This power is independent of either time T, or load Q_{Lt}.

Once we know the similarity criteria for economical alternatives, we may re-write Eq. (3.119) in criterion form (Fig. 3.17) as:

$$E_*=2Q_{c*}-Q_{c*}^2 \tag{3.121}$$

where

$$E_*=E/E_{\max}$$

$$Q_{c*}=Q_c/Q_{c.e}$$

It follows from the criterion equation that the maximum economy, $E_*=1$, is achieved at $Q=1$. Any departure from $Q_{c.e}$, such as under-compensation, $Q_{c*}<1$, or overcompensation, $Q_{c*}>1$, would lead to a

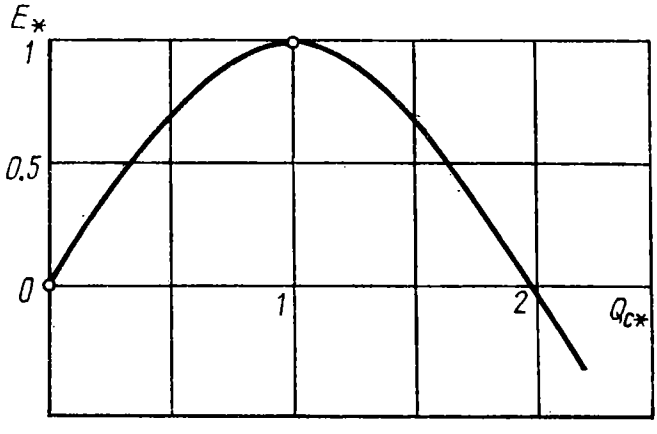


Fig. 3.17

reduction in economy. It is relevant to add that with more than two-fold overcompensation the economy becomes negative, that is, the cost rises. This conclusion holds for any segments of an electric network, because it has been derived from the criterion equation, with the constants A_i unknown.

If, on the other hand, the constants A_i are known, Eq. (3.120) may be used to determine the economical capacity of reactive-power compensators

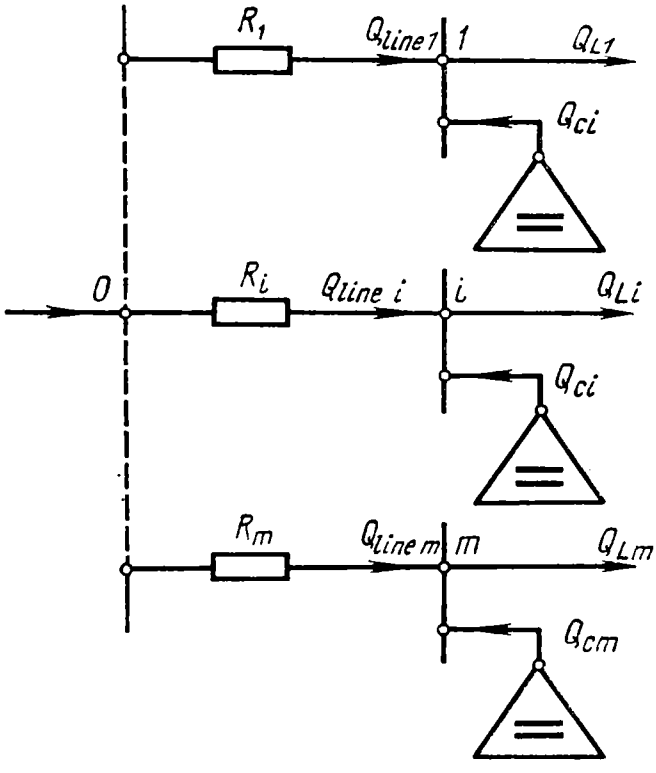


Fig. 3.18

to be installed on a power transmission line. In fact, it will then be possible to select and site reactive-power compensators for fairly complex networks.

To begin with, we shall consider a radial network (Fig. 3.18). We assume that there are m consumers supplied by a single source O and that there is a reactive power compensator of capacity Q_{ci} at each consumer's.

If we assume that the loads are mutually independent, we may use Eqs. (3.120) to determine the reactive powers $Q_{line\,ie}$ which can advantageously be transmitted to the receiving end of the line:

$$V_i^2(c_{ci}-c_0)/2b = R_i Q_{line\,ie}, \qquad i=\overline{1,m}$$

These equations can be written in contracted matrix form as

$$\mathbf{V}=\mathbf{R}\mathbf{Q}_{line\,e} \tag{3.122}$$

Since the resistance matrix is non-singular, we may use Eq. (3.122) to find that

$$\mathbf{Q}_{line\,e}=\mathbf{R}^{-1}\mathbf{V} \tag{3.123}$$

where

$$\mathbf{V}=\left\|\begin{matrix} V_1^2\,\frac{c_{c1}-c_0}{2b} \\ \vdots \\ V_m^2\,\frac{c_{cm}-c_0}{2b}\end{matrix}\right\| \qquad \mathbf{R}=\left\|\begin{matrix} R_1 & \dots \\ \dots & R_i \dots \\ & \dots\ R_m\end{matrix}\right\| \qquad \mathbf{Q}_{line\,e}=\left\|\begin{matrix} Q_{line\,e} \\ \vdots \\ Q_{line\,ie} \\ \vdots \\ Q_{line\,me}\end{matrix}\right\|$$

Now we shall consider a bulk supply line (Fig. 3.19), assuming that there is a reactive power compensator at each load.

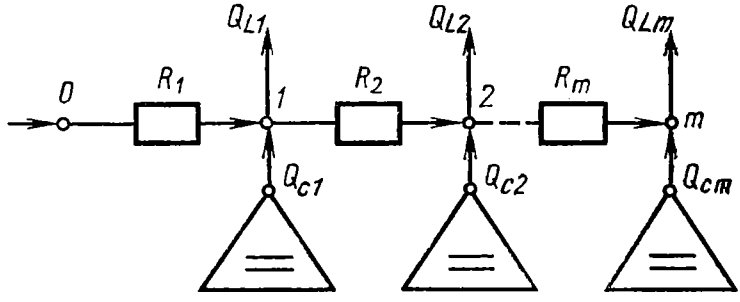


Fig. 3.19

If there were no compensators installed, the cost of reactive power transmission to the consumers would be

$$C_1=\sum_{i=1}^{m} R_i\left(\sum_{j=i}^{m} Q_{Lj}\right)^2(bT_i/V^2)+c_0\sum_{i=1}^{m} Q_{Li}T_i$$

With reactive power compensators installed, the cost becomes

$$C_2=\sum_{i=1}^{m} R_i\left[\sum_{j=i}^{m}(Q_{Lj}-Q_{cj})\right]^2\frac{bT_i}{V^2}+c_0\sum_{i=1}^{m}(Q_{Li}-Q_{ci})T_i+\sum_{i=1}^{m} c_{ci}Q_{ci}T_i$$

The economy due to the installation of a reactive power compensator is

$$E=C_1-C_2\,.$$

Differentiating the last equation with respect to Q_{ci} yields the following system of linear equations

$$R_{01}Q_{line\,1e}+R_{02}Q_{line\,2e}+\dots+R_{01}Q_{line\,me}=V_1^2\frac{c_1-c_0}{2b}$$

$$R_{01}Q_{line\,1e}+R_{02}Q_{line\,2e}+\dots+R_{02}Q_{line\,me}=V_2^2\frac{c_2-c_0}{2b}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$R_{01}Q_{line\,1e}+R_{02}Q_{line\,2e}+\dots+R_{0m}Q_{line\,me}=V_m^2\frac{c_m-c_0}{2b}$$

where

$$R_{0i}=\sum_{j=i}^{m} R_j$$

The above system may be written in matrix form, Eq. (3.122), and solved to obtain a solution in the form of Eq. (3.123).

It can be shown that for a ramified network supplied from one side only and of an arbitrary configuration (other than closed), the economical capacity of a reactive power compensator will be decided by Eqs. (3.123) and by the reactive power balance equations. Relatively complex networks should preferably be solved with the aid of a digital computer.

Consider an example in which the design equation contains negative constants. Such equations crop up in studies concerned with the effect that

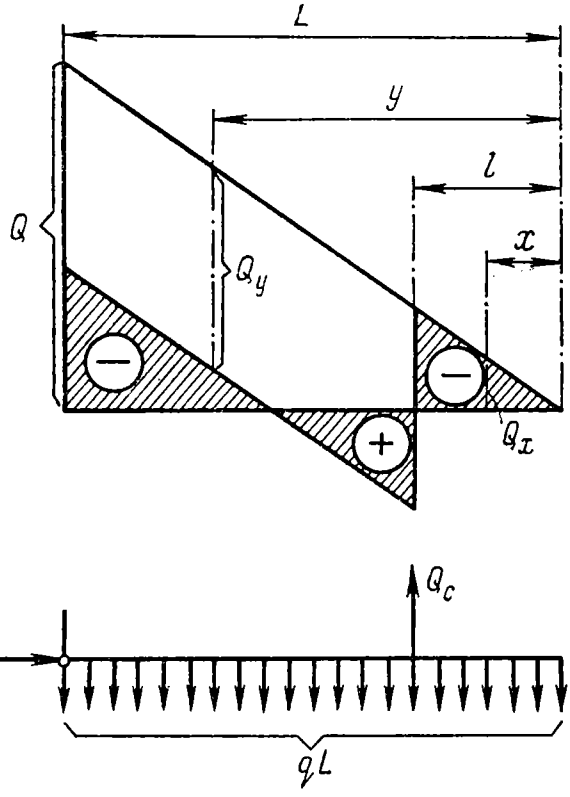


Fig. 3.20

the point of connection and the capacity of an unregulated reactive power compensator may have on transmission losses. Assume that reactive load is distributed along the line uniformly (Fig. 3.20) and that the load curve of all the recipients varies in time phase identically. Examples of such networks are the bus systems of factories and other industrial enterprises. The effectiveness of reactive power compensation can be assessed in terms of

the reduction in transmission loss which, subject to the assumptions made, will be

$$\Delta P=\left(\int_0^l Q_x^2\,dx+\int_l^L Q_y^2\,dy\right) r_0/V^2$$

Substituting in the above equation the reactive powers for the various sections of the line,

$$Q_x=Qx/L$$
$$Q_y=Qy/L-Q_c$$

integrating, and carrying out appropriate manipulations, we get

$$\Delta P=\frac{q^2 r_0 L^3}{3V^2}\left[(1-Q_*)^3+(Q_*-l_*)^2+l^3\right] \tag{3.124}$$

Here, q = linear density of reactive load
r_0 = resistivity of the line
L = length of the line;
$Q_*=Q_c/qL$ = relative capacity of the compensator
$l_*=l/L$ = ratio of line length between the end and the connection of the compensator to total line length
V = voltage of the network to which the line is connected

Equation (3.124) provides a means for technical and economic analysis of the problem stated earlier.

If we know the installed capacity of the reactive power compensator in per-unit, Q_*, the point of connection for the compensator, advantageous from an economical point of view, can be determined from the equation derived by equating the derivative to zero:

$$d\Delta P/dl_*=0$$

whence

$$l_{*e}=Q_*/2 \tag{3.125}$$

As is seen, the point of connection for a reactive power compensator depends on its capacity. Substituting Eq. (3.125) in Eq. (3.124), the power lost in the line is defined as

$$\Delta P_{qe}=\frac{q^2 r_0 L^3}{3V^2}\left[(1-Q_*)^3+Q_*^3/4\right] \tag{3.126}$$

If we know the location of the compensator on the line, l_*, its economical capacity can be defined from the equation

$$d\Delta P/dQ_*=0$$

whence

$$Q_{*e}=(1+l_*)/2 \tag{3.127}$$

In this case, the power lost in the line will be

$$\Delta P_{le}=\frac{q^2 r_0 L^3}{3V^2}\left[1-\frac{3}{4}(1+l_*)^2(1-l_*)\right] \tag{3.128}$$

Equations (3.125) and (3.126) through (3.128) define either the economical capacity of a reactive power compensator or an economical point of connection for the device, and the associated transmission power losses if the point of connection is specified in advance. If both indicators are to be optimized at the same time, then, by solving Eqs. (3.125) and (3.127) simultaneously, we get

$$\begin{aligned} l_{*e}&=1/3 \\ Q_{*e}&=2/3 \end{aligned} \tag{3.129}$$

By substituting these values in the design equation, Eq. (3.124), we can determine the minimum likely power lost in the line, under the assumptions made

$$\Delta P_e=q^2r_0L^3/(27V^2) \tag{3.130}$$

Taking this value as the basic one and dividing Eq. (3.124) into it leads to a criterion equation

$$\Delta P_*=9[(1-Q_*)^3+(Q_*-l_*)^3+l_*^3] \tag{3.131}$$

from which we can determine the relative increase in the power lost in the line for the point of connection and device capacity other than economical [see Eq. (3.129)]. For example, if no reactive power compensator is provided, then $Q_*=l_*=0$ and $\Delta P_*=9$. So, in lines with a uniformly distributed load, reactive power compensation by a single unregulated device can reduce power losses caused by the flow of reactive power over the line by a factor of nine. This level of effectiveness applies to all lines with a uniformly distributed load without exception, because it has been derived from analysis of the criterion equation, Eq. (3.131).

If either the point of connection or the capacity of a reactive power compensator need only be optimized, the gain accruing from such a device can be assessed from the criterion equations obtained by dividing Eqs. (3.126) and (3.128) into Eq. (3.130):

$$\begin{aligned} \Delta P_{q*}&=9\left[(1-Q_*)^3+\frac{1}{4}Q_*^3\right] \\ \Delta P_{l*}&=9\left[1-\frac{3}{4}(1+l_*)^2(1-l_*)\right] \end{aligned} \tag{3.132}$$

From analysis of criterion equations (3.132) shown in graphical form in Figs. 3.21 and 3.22 it is seen that optimization of only the point of installation or of only the capacity of a reactive power compensator might result in an incomplete utilization of its capability for reducing power losses in the line. It is more economically advantageous to optimize both factors at the same time.

Consider the selection of peak-load stations described by equations with negative generalized constants.

Suppose that it is planned to build two electric stations, one to supply the base and other the peak load of an area (Fig. 3.23*a*), with duration

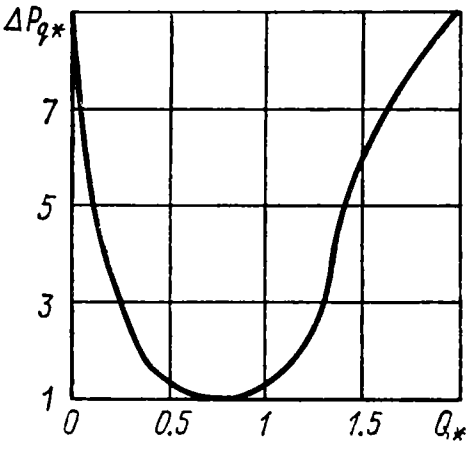


Fig. 3.21

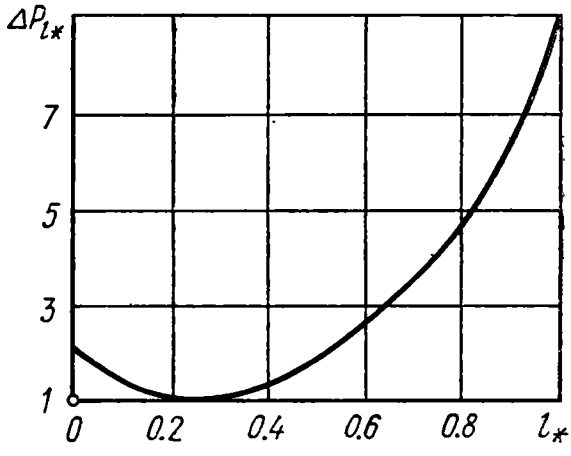


Fig. 3.22

curves as shown in Fig. 3.23*b*. The total capacity of the stations is equal to the maximum load power

$$P_{\max}=P_b+P_p \tag{3.133}$$

and the total energy output is equal to that of the loads

$$A_{\text{out}}=A_b+A_p \tag{3.134}$$

The overall discounted cost of power generation by both stations is

$$C=C_b+C_p \tag{3.135}$$

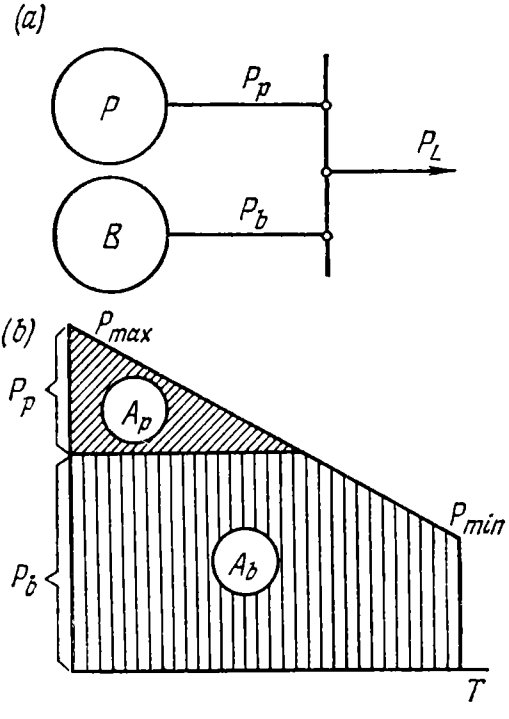


Fig. 3.23

To make the problem more concrete, we assume that the specific cost items are known and that

$$a_b>a_p, \qquad b_b<b_p$$

where a = specific cost of construction

b = specific cost of energy production

From these conditions it follows that station P should operate in the peak duty, and station B in the base duty. The power of the peak station will be the parameter to be optimized. Then, using the power balance equation, Eq. (3.133), and the energy balance equation, Eq. (3.134), the total cost, Eq. (3.135), may be defined as

$$C=C_{0b}+a_bP_{\max}+b_bA_L$$
$$+C_{0p}-(a_b-a_p)P_p+(b_p-b_b)A_p$$

The first three terms determine the costs

$$C_b = C_{0b} + a_b P_L + b_b A_L$$

that would result if no peak station were to be built and all load were to be supplied by the base station alone. The last three terms

$$\Delta C = C_{0p} - (a_b - a_p)P_p + (b_p - b_b)A_p$$

define the change in the costs due to the construction of a peak station. If we note that the energy generated by the peak station is

$$A_p = P_p^2 T / [2(P_{\max} - P_{\min})] \tag{3.136}$$

then the change in the costs due to the construction of the peak station would be

$$\Delta C = C_{0b} - (a_b - a_p)P_p - \frac{(b_p - b_b)P_p^2}{2(P_{\max} - P_{\min})} T \tag{3.137}$$

The last equation shows that the construction of a peak load station would increase the overall costs because of the presence of a constant term, C_{0p}, and the difference in specific costs ($b_b < b_p$). However, increasing the peak power would cause the overall costs to decrease because of the difference in the relevant specific costs ($a_b > a_p$).

The last two terms in Eq. (3.137), which are dependent on the capacity of the peak station, define the economy, E, which would arise from the construction of that station:

$$E = (a_b - a_p)P_p - \frac{(b_p - b_b)P_p^2}{2(P_{\max} - P_{\min})} T \tag{3.138}$$

The economical capacity of the peak station can be defined from the condition $dE/dP_p = 0$:

$$P_{pe} = (a_b - a_p)(P_{\max} - P_{\min}) / [(b_p - b_b)T] \tag{3.139}$$

Hence, the maximum economy would be

$$E_{\max} = (a_b - a_p)^2 (P_{\max} - P_{\min}) / [2(b_p - b_b)T] \tag{3.140}$$

Thus, the construction of a peak station would cause the total costs to change by

$$\Delta C = C_{0p} - (a_b - a_p)^2 (P_{\max} - P_{\min}) / [2(b_p - b_b)T] \tag{3.141}$$

It is seen from Eq. (3.139) that the economical capacity of the peak station would be proportional to the peak load power, $(P_{\max} - P_{\min})$, the difference in construction costs $(a_b - a_p)$, inversely proportional to the time interval under study, T, and inversely proportional to the difference in specific generation costs, $(b_p - b_b)$. If the specific construction cost were the same for both stations, $a_b = a_p$, no advantage would be gained from building a peak station, because it would only add to the fixed cost term, C_{0p}. Nor would it be advantageous to build a peak station if there were no peak term in the load,

$$P_{\max} - P_{\min} = 0$$

If both stations have the same specific generation costs, ($b_p = b_b$), it would not be economically advantageous to build a peak station ($P_{p0} = \infty$), as this would increase the total construction cost.

When the design data included in Eqs. (3.139) and (3.141) are known in advance, it is possible to determine the economical capacity of the peak station and the economy to be gained from its construction. However, in design work it is often necessary to know how economy may be affected, should some technical limitation cause the capacity of the peak station to be different from its economical value. We shall carry out this analysis, using an appropriate criterion equation.

In this equation, the basic quantities will be the economical capacity of the peak station and the economy to be gained, as defined by Eqs. (3.139) and (3.140). In criterion form, Eq. (3.139) may be written as

$$\Delta C_* = C_{0p*} - 2P_{p*} + P_{p*}^2$$

In Fig. 3.24 the criterion equation is plotted for several values of C_{0p*}. It would be economically advantageous to build a peak station only when

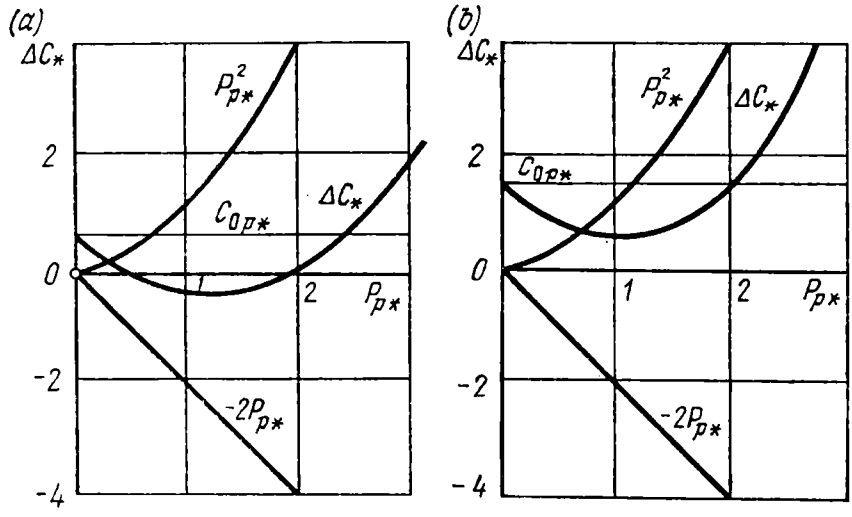


Fig. 3.24

the constant term in the construction cost of the peak station is less than the maximum likely economy:

$$C_{0p} < E_{\max}$$

or, in per-unit

$$C_{0p*} < 1$$

Then the limiting value of peak capacity can be determined from the condition

$$\Delta C_* = C_{0p*} - 2P_{p*} + P_{p*}^2 = 0$$

as

$$P_{p*1,2}=1\pm\sqrt{1-C_{0p*}}$$

It would not be advantageous to build a peak station with a capacity.

$$P_{p*1}<1-\sqrt{1-C_{0p*}}$$

or

$$P_{p*2}>1+\sqrt{1-C_{0p*}}$$

because the ensuing economy would not be sufficient to make up for the fixed term in the construction cost of the peak station. Hence, the capacity of the peak station should be chosen to lie in the range

$$P_{p*1}\leqslant P_{p*}\leqslant P_{p*2}$$

If a peak station is warranted, its power range will be

$$\Delta P_{p*}=P_{p*2}-P_{p*1}=2\sqrt{1-C_{0p*}}$$

The examples taken up in this section indicate that criterion analysis is likewise applicable to discounted cost equations containing negative generalized constants, A_i.

3.21. Criterion Analysis Covering the Reliability of Service

We shall show the usability of criterion analysis in cases where design data are probabilistic in nature, using the reliability of service as an example. A major road-block in optimizing electricity supply systems is the lack of a sufficiently objective estimate for the cost of loss to the national economy because of interruptions in service. This is why it is important to have criterion (dimensionless) relations that would allow analysis to be carried out on incomplete design data.

We shall consider how the reliability of a line can be enhanced through redundancy, that is, by simultaneously using m parallel circuits. Obviously, an increase in m leads to greater investment costs, while energy loss and loss due to interruption in service are reduced. The crux of the problem is to select an economical number of circuits to be used and, as a consequence, an economical reliability level of electricity supply.

However, it would be wrong to optimize only one parameter m, because a line is characterized by many interrelated parameters. To simplify computational work, we assume that, apart from m, we seek to optimize the conductor cross-sectional area F and the line voltage V. Then the design equation may be written as

$$C=A_V V_m^{\beta}+A_F Fm+A_s V^{-2}F^{-1}m^{-1}+A_m m^{-\alpha} \qquad (3.142)$$

For overhead power transmission lines, the exponents may be set equal to $\beta=1/2$ and $\alpha=2$.

The dimension matrix and its inverse for Eq. (3.142) are

$$\alpha=\left\|\begin{array}{cccc} m & V & F & C \\ \hline 1 & 1/2 & 0 & -1 \\ 1 & 0 & 1 & -1 \\ -1 & -2 & -1 & -1 \\ -2 & 0 & 0 & -1 \end{array}\right\| \qquad \alpha^{-1}=\left\|\begin{array}{c|cccc} m & 4/16 & 1/16 & 1/16 & -6/16 \\ V & 8/16 & -6/16 & -6/16 & 4/16 \\ F & -12/16 & 13/16 & -3/16 & 2/16 \\ C & -8/16 & -2/16 & -2/16 & -4/16 \end{array}\right\| \tag{3.143}$$

Hence, the similarity criteria are

$$\pi_{1e}=1/2, \quad \pi_{2e}=\pi_{3e}=1/8, \quad \pi_{4e}=1/4$$

and the equations defining an economical alternative can be written as

$$\left.\begin{aligned} m_e&=\left(\frac{4A_m^6}{A_V^4A_FA_s}\right)^{1/16}, & V_e&=2\left(\frac{4A_F^3A_s^3}{A_V^4A_m^2}\right)^{1/8} \\ F_e&=\frac{1}{2}\left(\frac{A_V^{12}A_s^3}{256A_F^{13}A_m^2}\right)^{1/16}, & C_e&=4\left(\frac{A_V^4A_FA_sA_m^2}{4}\right)^{1/8} \end{aligned}\right\} \tag{3.144}$$

The criterion equation

$$C_*=\frac{1}{2}V_*^{1/2}m_*+\frac{1}{8}F_*m_*+\frac{1}{8}V_*^{-2}F_*^{-1}m_*^{-1}+\frac{1}{4}m_*^{-2}$$

gives an estimate of economical stability. For example, should any of the terms deviate from its economical value in the range 0.75 to 1.5, the increase in the respective cost will be $C_{F*}=1.01$ to 1.02, $C_{V*}=1.03$ to 1.04, and $C_{m*}=1.08$ to 1.13.

Now we shall solve the dual problem: we assume to know in advance the limit of increase in the cost, say $C_*=1.05$, and determine the likely departure from the economical value for each term, letting the remaining ones remain equal to their economical values:

$$\begin{array}{lll} m_*=1, & V_*=1, & F_*=0.37 \text{ to } 1.63 \\ m_*=1, & F_*=1, & V_*=0.6 \text{ to } 1.4 \\ V_*=1, & F_*=1, & m_*=0.76 \text{ to } 1.24 \end{array}$$

Thus the allowable departure for conductor cross-sectional area is nearly three times as great as that for the number of circuits. In other words, the model under consideration is very stable towards cross-sectional area and has poor stability towards the number of circuits. Hence, in design work, special emphasis need be placed on economic study of the number of circuits to be installed.

Now we shall analyse the economic values of the various parameters and costs for sensitivity towards changes in the constant A_m. This constant relates to the cost of loss due to interruptions in service and is usually least precise of all. From Eq. (3.144) it follows that

$$\mathbf{m}_*=A_{m*}^{3/8}, \qquad V_*=A_{m*}^{-1/4}, \qquad F_*=A_{m*}^{-1/8}, \qquad C_*=A_{m*}^{1/4}$$

From analysis of the above relations, it may be concluded that the most sensitive and the least sensitive parameter towards inaccuracy in A_m is **m** and F, respectively. For example, should the cost of interruption in electricity supply be overestimated or underestimated by a factor of 1.5, the remaining parameters will depart from their economical values within the ranges defined below:

$$\mathbf{m}_* = 0.86 \text{ to } 1.17, \qquad V_* = 1.10 \text{ to } 0.91, \qquad F_* = 1.05 \text{ to } 0.95$$
$$C_* = 0.91 \text{ to } 1.10$$

If the estimate of the costs is to remain within the specified limit of accuracy (5%), the cost of loss due to interruption in electricity supply must be estimated accurate to within 20%. As is seen, no special precision is needed in estimating the cost of loss due to interruption in electricity supply. This can be borne out by an example.

Suppose a line is planned, 10 km long, to transmit 5 MW of electricity for a shoe-making factory. To begin with, we evaluate the generalized constants in Eq. (3.142):

$$A_V = 400 \text{ roubles/kW}^{1/2}, \quad A_F = 12 \text{ roubles/mm}^2,$$
$$A_s = 2.8 \times 10^8 \text{ kW}^2 \text{ roubles mm}^2, \quad A_m = 15 \times 10^3 \text{ roubles}$$

We shall compare four alternatives:

(1) the economical alternative defined by Eqs. (3.144);

(2) an optimal alternative which differs from the economical one because the voltage, cross-sectional area and number of circuits are limited each to a range of discrete standard values and the calculated values have to be rounded off to fit them;

(3) the economical alternatives defined by the model, Eq. (3.142), in which the last term is missing, which is the cost of loss due to interruption in electricity supply;

(4) an optimal alternative which differs from that immediately above in that the respective figures are rounded off to standard values.

The results are summarized in Table 3.2.

As is seen, neglect of the cost of loss due to interruptions in electricity supply results in a nearly two-fold increase in costs, although the economi-

Table 3.2

Comparison of Alternative Designs

Alternative	Parameters					
	V, kilovolts	F, mm²	m, circuits	j, A/mm²	C	
					roubles	%
1	41	96	2.3	0.35	11,700	100
2	35	95	2.0	0.42	12,160	104
3	49	70	1.0	0.92	20,060	172
4	35	95	1.0	0.97	20,210	175

cal current density remains within the limits recommended by appropriate codes. The costs can be radically cut down in the optimal alternative where the cost of loss due to interruption in electricity supply is determined, even though with low accuracy. This entails a two-fold increase in the number of circuits and a two-fold decrease in line current density.

3.22. Criterion Analysis of Equations Containing Exponential Functions

In any system of units, a measure (a standard) is customarily some quantity. However, it would be as legitimate to use a function as a unit of comparison, that is, to employ the units of a function in dimensional analysis.

For example, if we use a function, $f(x_0)$, as a standard or measure, the function $f(x)$ can be expressed in terms of this standard as

$$[f(x)]=f(x)/f(x_0) \quad \text{f.u.}$$

where f.u. stands for functional units.

Engineering widely uses this system of functional units, as it is more convenient and graphic. For example, it is usual to refer to some vessel (a glass, pail, barrel, etc.) as a measure of volume. In the theory of information, the unit is the bit (or binit) which is the logarithm of the number 2 taken to the base 2,

$$\log_2 2=1 \text{ bit (binit)}$$

Similarly, in acoustics, electrical engineering and radio engineering, use is made of functional units such as the bell, the neper, and the decilog which are functions of the logarithm of the ratio of two quantities.

Functional units extend the applicability of the theory of similarity to technico-economic and purely physical problems. Also, functional units raise the theory of similarity to the status of a theory of isomorphism.

As an example illustrating the use of functional units in criterion analysis, we shall examine an equation containing exponential functions:

$$C=A_1 \exp(\alpha_1 x)+A_2 \exp(-\alpha_2 x) \tag{3.145}$$

Such equations arise in economical analysis of reliability to be provided in multiply reclosable networks.

In Eq. (3.145), the first term represents the operating costs and the second the cost of improvements in the system reliability to be provided during its design or manufacture. The variable being optimized, x, is the ratio of the system's operating time to the mean time between failures; the constants A_1, A_2, and α_1, α_2 take values specific for specific systems.

Equation (3.145) is a special case of the more general equation

$$C=\sum_{j=1}^{m} A_i \prod_{j=1}^{n} a^{x_{ij}x_j} \tag{3.146}$$

By change of variable,

$$z_j = a^{x_j}$$

we obtain the following equation

$$C = \sum_{i=1}^{m} A_i \prod_{j=1}^{n} z_j^{\alpha_{ij}} \tag{3.147}$$

for which the technique of criterion analysis is well-developed. The models, Eqs. (3.146) and (3.147), will have canonical form only when the determinant of the matrix α is non-zero. Then the similarity criteria of economical alternatives will be independent of the design data, or constants, A_i. The economical values of the equation variables can be defined by equations of the form

$$x_{je} = \sum_{i=1}^{n+1} \frac{\Delta_{ji}}{\Delta} \log \frac{\pi_{ie}}{A_i \log a}, \qquad j = \left| \begin{matrix} n \\ 1 \end{matrix} \right. \tag{3.148}$$

and the minimum cost as

$$C_e = \prod_{i=1}^{n+1} \left(\frac{\pi_{ie}}{A_i \log a} \right)^{\pi_{ie}} \tag{3.149}$$

Equations (3.148) and (3.149) are similar to Eqs. (3.61) and (3.62), although the equations for x_{je} and C_e are different. The criterion equation

$$C_* = \sum_{i=1}^{m} \pi_{ie} \prod_{j=1}^{n} a_*^{\alpha_{ij} x_j} \tag{3.150}$$

will likewise be different from Eq. (3.54). In this equation the relative (per-unit) cost is

$$C_* = C/C_e$$

Instead of the economical values of the variables, however, use is made of relative functional units

$$a_*^{\alpha_{ij} \Delta x_j} = a^{\alpha_{ij} x_j} / a^{\alpha_{ij} x_{je}}$$

where the functional reference unit is the value of the respective function occurring when the variables have economical values.

In spite of some differences, Eqs. (3.148) through (3.150) lend themselves to criterion analysis to the same extent as the cost equations [see Eq. (3.49)].

For example, at $\alpha_1 = 1$ and $\alpha_2 = 1.25$, Eq. (3.145) written in criterion form

$$C_* = 5e_*^{x}/9 + 4e_*^{-1.25x}/9 \tag{3.151}$$

is of generalized character. From analysis of the criterion equation it is seen that changes of $\pm 28\%$ in the reliability parameter entail an increase of not more than 5% in the costs ($C_* = 1.05$). Therefore, this model may be taken as being sufficiently stable economically.

3.23. Basic Outline of Criterion Programming

The preceding sections have dealt with the criterion analysis of technico-economic equations having canonical form. Main emphasis has been placed on the so-called economical alternative. This has been done for several reasons. For one thing, there should be a single common basic alternative against which all other alternatives can be compared. For another, the techniques used to identify a basic alternative must yield univalued results. From these two view-points, the economical alternative is most advantageous because, given an equation in canonical form, it can be found by a technique which, for all of its specific features, is conventional and relies on equating to zero the first derivative of the discounted costs with respect to the variable being optimized.

With the economical alternative as a fixed and unchangeable reference, criterion analysis can be used to tackle problems for which an extremely limited information may sometimes be available.

For all computational conveniences, however, it is important to remember that in engineering work one has often to sacrifice the economical alternative because of existing limitations. Therefore, it is recommended to realize an optimal alternative which takes proper account of these limitations. As a rule, it involves greater discounted costs in comparison with the economical alternative. It takes the engineer's skills to select the most attractive alternative design from among a multitude of practically equal ones.

The solution of optimization problems by techniques of criterion analysis, with due account of technical limitations, constitutes the essence of criterion programming. In the general form, the primary task of criterion programming may be stated as follows. Minimize the function

$$u(x), \qquad x=(x_1, x_2, \ldots, x_n) \tag{3.152}$$

subject to

$$x_1>0, \qquad x_2>0, \ldots, x_n>0 \tag{3.153}$$

and

$$q_1(x)\leqslant 1, \qquad q_2(x)\leqslant 1, \ldots, q_p(x)\leqslant 1 \tag{3.154}$$

It is assumed that the optimality criterion, $u(x)$, and the constraints, $q(x)$, are expressed by generalized polynomials of the form

$$u(x)=\sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_j^{\alpha_{ij}} \tag{3.155}$$

It is usual to choose the discounted costs C as the optimality criterion, although, in the general case, this may be any other indicator, such as weight, volume, energy losses, etc. This is why we shall use u, and not C, as the criterion of optimality.

A distinction of criterion programming is that principal emphasis is placed on the terms $u_i(x)$ of the optimality criterion, Eq. (3.155):

$$\left.\begin{aligned} u_i&=A_i \prod_{j=1}^{n} x_j^{\alpha_{ij}}, \quad i=\overline{1\,|\,m} \\ u_1+u_2+\ldots+u_m&=u \end{aligned}\right\} \tag{3.156}$$

While in the traditional methods of mathematical programming, attention is focused on the variables being optimized, x_j, criterion programming is centred around the relative contributions by the various components of the optimality criterion, that is, the technico-economic similarity criteria

$$\left.\begin{aligned} \pi_i &= \frac{u_i}{u} = \frac{A_i}{u} \prod_{j=1}^{n} x_j^{\alpha_{ij}} \\ \pi_1 &+ \pi_2 + \ldots + \pi_m = 1 \end{aligned}\right\} \tag{3.157}$$

The fact that similarity criteria are pivotal in mathematical programming is reflected in the name of the method — criterion programming.

3.24. Geometric Inequality and Minimization Problems

Criterion programming widely uses the so-called geometric inequality in which the arithmetic mean of numbers is constrained never to be less than their geometric mean.

The simplest geometric inequality has the form

$$u_1/2 + u_2/2 \geqslant u_1^{1/2} u_2^{1/2} \tag{3.158}$$

where u_1 and u_2 may be any non-negative numbers.

To prove Eq. (3.158), it is to be noted that

$$(u_1 - u_2)^2 \geqslant 0$$

or

$$u_1^2 - 2u_1u_2 + u_2^2 \geqslant 0$$

whence

$$u_1^2 + 2u_1u_2 + u_2^2 \geqslant 4u_1u_2$$

Extracting the square root from the last equation proves the inequality, Eq. (3.158). It may be added that the inequality (3.158) becomes an equality when and only when $u_1 = u_2$.

As an example of the inequality, Eq. (3.158), consider the arithmetic mean of four non-negative numbers u_1, u_2, u_3 and u_4. By Eq. (3.158),

$$u_1/4 + u_2/4 + u_3/4 + u_4/4 \geqslant \left(\frac{u_1 + u_2}{2}\right)^{1/2} \left(\frac{u_3 + u_4}{2}\right)^{1/2}$$

Using Eq. (3.158) once more, we get

$$u_1/4 + u_2/4 + u_3/4 + u_4/4 \geqslant u_1^{1/4} u_2^{1/4} u_3^{1/4} u_4^{1/4} \tag{3.159}$$

Inequality (3.159) turns to an equality only when

$$u_1 = u_2 = u_3 = u_4$$

Setting $u_1 = u_3 = u_4 \neq u_2$, the inequality may be re-written as

$$u_1/4 + 3u_2/4 \geqslant u_1^{1/4} u_2^{3/4} \tag{3.160}$$

The left-hand side of the above expression is the arithmetic mean of the numbers u_1 and u_2 with weights 1/4 and 3/4. The right-hand side of the

expression is the weighted geometric mean. Introducing appropriate weights and setting Eq. (3.158) against Eq. (3.160), we may write in the generalized form

$$\pi_1 u_1 + \pi_2 u_2 \geqslant u_1^{\pi_1} u_2^{\pi_2} \tag{3.161}$$

where the weights are chosen such that their sum is unity:

$$\pi_1 + \pi_2 = 1 \tag{3.162}$$

Two numbers, u_1 and u_2, are compared in Eq. (3.161). If we examine m numbers $(u_1, u_2, \ldots, u_m)$, a similar proof can be constructed for the inequality

$$\pi_1 u_1 + \pi_2 u_2 + \ldots + \pi_m u_m \geqslant u_1^{\pi_1} u_2^{\pi_2} \ldots u_m^{\pi_m} \tag{3.163}$$

which is a logical generalization of Eq. (3.161). The weights are subject to the normality condition:

$$\sum_{i=1}^{m} \pi_i = 1 \tag{3.164}$$

In a more compact form, Eq. (3.163) can be written as

$$\sum_{i=1}^{m} \pi_i u_i \geqslant \prod_{i=1}^{m} u_i^{\pi_i} \tag{3.165a}$$

However, Eq. (3.165*a*) may be written in another way. To prove this, we introduce new variables, say

$$V_i = \pi_i u_i$$

and write the original equation in terms of the new variables:

$$\sum_{i=1}^{m} V_i \geqslant \prod_{i=1}^{m} \left(\frac{V_i}{\pi_i}\right)^{\pi_i} \tag{3.165b}$$

The above relation is very general and will not therefore change if we change the literals:

$$\sum_{i=1}^{m} u_i \geqslant \prod_{i=1}^{m} \left(\frac{u_i}{\pi_i}\right)^{\pi_i} \tag{3.165c}$$

Thus, the original geometric inequality may be written in the form of both Eq. (3.165*a*) and Eq. (3.165*c*). These relations form the basis for a theory of criterion programming and, above all, for finding the lower bound of the generalized polynomials (posynomials). As will be recalled, the number $\mu > 0$ is called the lower bound on a function $u(x)$ if $u(x) \geqslant \mu$ for all changes in the argument within the domain D, that is, for $x \in D$.

Before we proceed to describing a general technique for finding the minimum of generalized polynomials by means of geometric inequalities, it will be instructive to consider a few examples. As the first example, let there be a polynomial of the form

$$u(x) = 4x + 1/x \tag{3.166}$$

a plot of which is shown in Fig. 3.25. As is seen, an increase in x brings about an increase in the first term and a decrease in the second; that is, x gives rise to a competition. The function $u(x)=4$ reaches a minimum at $x_e=1/2$.

To determine the extremum analytically, we equate the first derivative of $u(x)$ to zero:

$$du(x)/dx=4-1/x_e^2=0$$

whence,

$$x_e=1/2$$

To determine the minimum value of $u(x_e)$, we substitute the value of x_e in $u(x)$. As a result, $u(x_e)=4$.

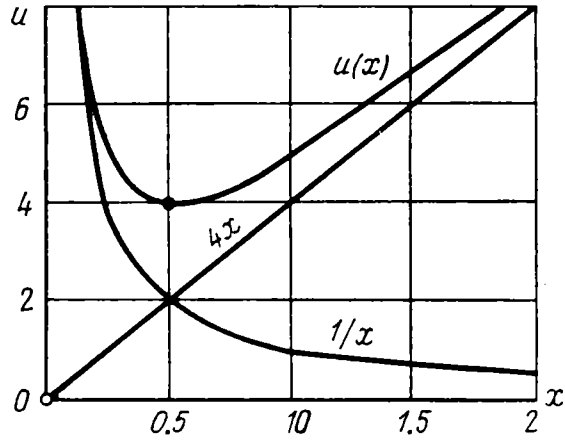


Fig. 3.25

The economical alternative of Eq. (3.166) can be found by use of Eq. (3.168). For this purpose we divide Eq. (3.166) by two:

$$u(x)/2=4x/2+1/(2x)$$

By Eq. (3.158), we get

$$u(x)/2=4x/2+1/(2x)\geqslant(4x)^{1/2}(1/x)^{1/2}=2$$

Therefore, for any positive value of x the lower bound is

$$u(x)\geqslant 4$$

In the plot, the bound is at $x_e=0.5$.

It is to be noted that in determining an economical alternative by the classical method of differential calculus, x_e is determined first and $u(x_e)$ next. With geometric inequalities, the procedure is the other way around: $u(x_e)$ is found first and x_e next. This is a fundamental difference between criterion programming and the classical methods used in the solution of optimization problems.

As a second example, consider the function, Eq. (3.70), using specific constants:

$$u(x)=2x_1+4x_2+1/(4x_1^2x_2)$$

To determine the lower bound on $u(x)$, we shall use the geometric inequality, Eq. (3.165c)*:

$$u_1+u_2+u_3\geq\left(\frac{u_1}{1/2}\right)^{1/2}\left(\frac{u_2}{1/4}\right)^{1/4}\left(\frac{u_3}{1/4}\right)^{1/4}$$

Then,

$$u(x)\geq\left(\frac{2x_1}{1/2}\right)^{1/2}\left(\frac{4x_2}{1/4}\right)^{1/4}\left(\frac{1}{4x_1^2x_2\times 0.25}\right)^{1/4}=4$$

Thus the lower bound on the function of two variables in question is 4.

Consider the polynomial, Eq. (3.77), written as

$$u(x)=x_1^{-2}+2x_1x_2^{-2}+4x_2$$

The original geometric inequality may be written as

$$u_1/7+2u_2/7+4u_3/7\geq(u_1)^{1/7}(u_2)^{2/7}(u_3)^{4/7}$$

Hence,

$$u(x)\geq\left(\frac{1}{(1/7)x_1^2}\right)^{1/7}\left(\frac{2x_1}{(2/7)x_2^2}\right)^{2/7}\left(\frac{4x_2}{4/7}\right)=7$$

Thus, the sought minimum is 7, and it has been evaluated without computing the parameters which minimize the original function.

From the examples examined above, it is seen that geometric inequalities can be used to evaluate the lower bounds on posynomials. However, it is unclear how the specific form of the inequality can be determined, why the lower bound is the minimum value of a particular posynomial, and how the parameters minimizing a given function can be evaluated. To answer these questions, we shall consider a technique for finding appropriate inequalities on a systematic basis.

3.25. Dual Functions

Consider the original geometric inequality in the form of Eq. (3.165c). The left-hand side of this inequality is the primal function, $u(x)$. The right-hand side of the same inequality is a multiplicative function which will further be referred to as its dual function and designated as

$$D(\pi)=\prod_{i=1}^{m}\left(\frac{u_i}{\pi_i}\right)^{\pi_i} \tag{3.167}$$

Inequality (3.165c) may be written in a more compact form as

$$u(x)\geq D(\pi) \tag{3.168}$$

* The values of π_i are given on page 195 [see Eq. (3.75a)].

If the function $u(x)$ is a generalized polynomial, Eq. (3.155), then, by substituting u_i from Eq. (3.156) in Eq. (3.168), we get

$$u(x) \geqslant \left(\frac{A_1}{\pi_1}\right)^{\pi_1} \ldots \left(\frac{A_m}{\pi_m}\right)^{\pi_m} x_1^{\Delta_1} \ldots x_n^{\Delta_n} \tag{3.169}$$

where the exponents are obtained as linear combinations of the exponents α_{ij} and defined by an equation of the form

$$\Delta_j = \sum_{i=1}^{m} \pi_i \alpha_{ij}, \qquad j = \overline{|}^{n}_{1} \tag{3.170}$$

If the similarity criteria, π_i, be chosen to satisfy the orthogonality condition, so that all exponents Δ_j remain equal to zero

$$\Delta_j = \sum_{i=1}^{m} \pi_i \alpha_{ij} = 0, \qquad j = \overline{|}^{n}_{1} \tag{3.171}$$

then Eq. (3.169) may be simplified as

$$u(x) \geqslant \left(\frac{A_1}{\pi_1}\right)^{\pi_1} \ldots \left(\frac{A_m}{\pi_m}\right)^{\pi_m} \tag{3.172}$$

We shall prove that in this case the minimum of the primal function $u(x)$ is the same as the maximum of the dual function $D(\pi)$.

We assume that at point $x_e(x_{1e}, x_{2e}, \ldots, x_{ne})$ with positive coordinates, $x_{je} > 0$, the posynomial $u(x)$ [see Eq. (1.155)] takes on a minimum value:

$$u(x_e) = \sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_{je}^{\alpha_{ij}} \tag{3.173}$$

Then the similarity criteria will be

$$\pi_{ie} = \frac{A_i}{u(x_e)} \prod_{j=1}^{n} x_{je}^{\alpha_{ij}}, \qquad i = \overline{|}^{m}_{1} \tag{3.174}$$

From mathematical analysis it is known that at an extremal point x_e the partial derivatives of a function with respect to each variable x_t vanish

$$\frac{\partial u(x)}{\partial x_t} = \sum_{i=1}^{m} \frac{\alpha_{it} A_i}{x_{te}} \prod_{j=1}^{n} x_{je}^{\alpha_{ij}} = 0, \qquad t = \overline{|}^{n}_{1} \tag{3.175}$$

By solving Eqs. (3.174) and (3.175), we obtain

$$\sum_{i=1}^{m} \alpha_{it} \pi_{ie} = 0, \qquad t = \overline{|}^{n}_{1} \tag{3.176}$$

Thus, the vector of similarity criteria $\pi_e(\pi_{1e}, \ldots, \pi_{me})$ satisfies the orthogonality condition. Therefore, Eq. (3.172) may be written as

$$u(x_e) \geqslant \left(\frac{A_1}{\pi_{1e}}\right)^{\pi_{1e}} \left(\frac{A_2}{\pi_{2e}}\right)^{\pi_{2e}} \ldots \left(\frac{A_m}{\pi_{me}}\right)^{\pi_{me}} \tag{3.177}$$

Besides, it is clear from Eqs. (3.173) and (3.174) that the similarity criterion vector satisfies the normality condition

$$\sum_{i=1}^{m} \pi_{ie} = 1 \tag{3.177a}$$

on the strength of which

$$[u(x_e)]^1 = [u(x_e)]^{\pi_{1e}} [u(x_e)]^{\pi_{2e}} \dots [u(x_e)]^{\pi_{me}}$$

By replacing $u(x_e)$ on the right-hand side in accordance with Eq. (3.174)

$$u_i(x_e) = \frac{A_i}{\pi_{ie}} \prod_{j=1}^{n} x_{je}^{\alpha_{ij}}$$

we get

$$u(x_e) = \left(\frac{A_1}{\pi_{1e}}\right)^{\pi_{1e}} \dots \left(\frac{A_m}{\pi_{me}}\right)^{\pi_{me}} x_{1e}^{\Delta_1} \dots x_{ne}^{\Delta_n}$$

Since, according to the orthogonality condition, Eq. (3.176), $\Delta_j = 0$, then

$$u(x_e) = \left(\frac{A_1}{\pi_{1e}}\right)^{\pi_{1e}} \dots \left(\frac{A_m}{\pi_{me}}\right)^{\pi_{me}} \tag{3.177b}$$

From comparison of Eqs. (3.173) and (3.177*b*) it is seen that at the minimum point the inequality gives way to an identity and so we may re-write the condition, Eq. (3.168), as

$$u(x) \geqslant \mu \geqslant D(\pi)$$

On the other hand, by Eq. (3.177*b*)

$$u(x_e) = D(\pi_e)$$

From the last two relations we may conclude that the minimum of the primal function is the same as the maximum of the dual function.

Example 3.5. Find the minimum of the function

$$u = 40/(x_1 x_2 x_3) + 40 x_2 x_3 + 20 x_1 x_3 + 10 x_1 x_2$$

In accordance with Eq. (3.177*b*) we may write

$$u(x_e) = \left(\frac{40}{\pi_{1e}}\right)^{\pi_{1e}} \left(\frac{40}{\pi_{2e}}\right)^{\pi_{2e}} \left(\frac{20}{\pi_{3e}}\right)^{\pi_{3e}} \left(\frac{10}{\pi_{4e}}\right)^{\pi_{4e}}$$

In order to evaluate four similarity criteria, we shall write three orthogonality conditions in accordance with Eq. (3.176) and one normality condition in accordance with Eq. (3.177*a*):

$$-\pi_{1e} + \pi_{3e} + \pi_{4e} = 0, \qquad -\pi_{1e} + \pi_{2e} + \pi_{4e} = 0$$

$$-\pi_{1e} + \pi_{2e} + \pi_{3e} = 0, \qquad \pi_{1e} + \pi_{2e} + \pi_{3e} + \pi_{4e} = 1$$

Solving these equations simultaneously yields:

$$\pi_{1e} = 2/5, \qquad \pi_{2e} = \pi_{3e} = \pi_{4e} = 1/5$$

Substituting the above similarity criteria in the dual function gives

$$u(x_e)=\left(\frac{40}{2/5}\right)^{2/5}\left(\frac{40}{1/5}\right)^{1/5}\left(\frac{20}{1/5}\right)^{1/5}\left(\frac{10}{1/5}\right)^{1/5}=100$$

Thus, criterion programming allows $u(x_e)$ to be evaluated without first computing the minimizing point x_e. This contrasts with the usual method of solving minimization problems based on the classical differential calculus. If it is necessary to evaluate x_e, use should be made of Eq. (3.174):

$$2\times 100/5=u_{1e}=40x_{1e}^{-1}x_{2e}^{-1}x_{3e}^{-1}, \qquad 1\times 100/5=u_{2e}=40x_{2e}x_{3e}$$
$$1\times 100/5=u_{3e}=20x_{1e}x_{3e}, \qquad 1\times 100/5=u_{4e}=10x_{1e}x_{2e}$$

Solving the above equations simultaneously yields:

$$x_{1e}=2, \qquad x_{2e}=1, \qquad x_{3e}=1/2$$

Thus, criterion programming makes it possible to evaluate all parameters of an economical alternative, $[x_e, u(x_e)]$, when no constraints are imposed on changes in the parameters. In fact, it may also be used when some functional constraints are imposed on changes in the optimizing variables. These constraints may be more general than those examined in Sec. 3.17. We shall demonstrate this capability on an example and shall then outline a general theory.

The function to be minimized is

$$u(x)=A_1/(x_1x_2x_3)+A_2x_2x_3 \tag{3.178}$$

subject to the constraint

$$A_3x_1x_3+A_4x_1x_2\leqslant 1 \tag{3.179}$$

To solve this problem, we should above all write the geometric inequality, Eq. (3.165*a*), in a more general form where all weights of the terms u_i, the similarity criteria π_i, are unnormalized. Let the unnormalized weights be $\Delta_1, \Delta_2, \ldots, \Delta_m$ and their sum be:

$$\lambda=\Delta_1+\ldots+\Delta_m$$

In order to normalize the weights, we divide the last equation into λ:

$$1=\Delta_1/\lambda+\ldots+\Delta_m/\lambda$$

The number Δ_i/λ represents the normalized weights of the terms in the original equation:

$$\Delta_i/\lambda=\pi_i, \qquad i=\overline{|}^{m}_{1}$$

If we replace the terms π_i in Eq. (3.165*c*) by the unnormalized weights Δ_i, we get

$$u_1+\ldots+u_m\geqslant\left(\frac{u_1}{\Delta_1}\right)^{\Delta_1/\lambda}\ldots\left(\frac{u_m}{\Delta_m}\right)^{\Delta_m/\lambda}\lambda$$

For convenience, this inequality may be written as

$$(u_1+\ldots+u_m)^{\lambda}\geqslant(u_1/\Delta_1)^{\Delta_1}\ldots(u_m/\Delta_m)^{\Delta_m}\lambda^{\lambda} \tag{3.180}$$

Now we re-state the original problem in accordance with the resultant inequality, Eq. (3.180).

Minimize the function, Eq. (3.178):

$$u(x)=u_1+u_2 \tag{3.181}$$

subject to the constraint, Eq. (3.179):

$$1\geqslant u_3+u_4 \tag{3.182}$$

In accordance with Eq. (3.180), Eqs. (3.181) and (3.182) may be re-written as

$$u(x)^{\lambda_0}\geqslant\left(\frac{u_1}{\Delta_1}\right)^{\Delta_1}\left(\frac{u_2}{\Delta_2}\right)^{\Delta_2}\lambda_0^{\lambda_0} \tag{3.183}$$

$$1^{\lambda_1}\geqslant\left(\frac{u_3}{\Delta_3}\right)^{\Delta_3}\left(\frac{u_4}{\Delta_4}\right)^{\Delta_4}\lambda_1^{\lambda_1} \tag{3.184}$$

where

$$\lambda_0=\Delta_1+\Delta_2, \qquad \lambda_1=\Delta_3+\Delta_4$$

Multiplying together the left- and right-hand sides of Eqs. (3.183) and (3.184) gives

$$u(x)^{\lambda_0}\geqslant\left(\frac{u_1}{\Delta_1}\right)^{\Delta_1}\left(\frac{u_2}{\Delta_2}\right)^{\Delta_2}\left(\frac{u_3}{\Delta_3}\right)^{\Delta_3}\left(\frac{u_4}{\Delta_4}\right)^{\Delta_4}\lambda_0^{\lambda_0}\lambda_1^{\lambda_1} \tag{3.185}$$

The above inequality, Eq. (3.185), remains valid for any choice of positive weights Δ_i. For simplicity, however, it is convenient to normalize λ_0 so that $\lambda_0=1$. Then

$$u(x)\geqslant\left(\frac{u_1}{\pi_1}\right)^{\pi_1}\left(\frac{u_2}{\pi_2}\right)^{\pi_2}\left(\frac{u_3}{\Delta_3}\right)^{\Delta_3}\left(\frac{u_4}{\Delta_4}\right)^{\Delta_4}\lambda_1^{\lambda_1}$$

Passing to the constants A_i, we obtain

$$u(x)\geqslant\left(\frac{A_1}{\pi_1}\right)^{\pi_1}\left(\frac{A_2}{\pi_2}\right)^{\pi_2}\left(\frac{A_3}{\Delta_3}\right)^{\Delta_3}\left(\frac{A_4}{\Delta_4}\right)^{\Delta_4}(\Delta_3+\Delta_4)^{\Delta_3+\Delta_4}x_1^{D_1}x_2^{D_2}x_3^{D_3}$$

where

$$D_1=-\pi_1+\Delta_3+\Delta_4, \qquad D_2=-\pi_1+\pi_2+\Delta_4, \qquad D_3=-\pi_1+\pi_2+\Delta_3$$

If all D_i were equal to zero, that is, if the terms π_1, π_2, π_3 and π_4 satisfied the orthogonality condition, then

$$u(x)\geqslant\left(\frac{A_1}{\pi_1}\right)^{\pi_1}\left(\frac{A_2}{\pi_2}\right)^{\pi_2}\left(\frac{A_3}{\Delta_3}\right)^{\Delta_3}\left(\frac{A_4}{\Delta_4}\right)^{\Delta_4}(\Delta_3+\Delta_4)^{\Delta_3+\Delta_4} \tag{3.186}$$

From the orthogonality and normality conditions,

$$-\pi_{10}+\Delta_{30}+\Delta_{40}=0, \qquad -\pi_{10}+\pi_{20}+\Delta_{40}=0$$

$$-\pi_{10}+\pi_{20}+\Delta_{30}=0, \qquad \pi_{10}+\pi_{20}=1$$

the similarity criteria are evaluated to be

$$\pi_{10}=2/3, \qquad \pi_{20}=\pi_{30}=\pi_{40}=1/3$$

Substituting the above values in Eq. (3.186) and setting, say, $A_1=40$, $A_2=40$, $A_3=1/2$ and $A_4=1/4$, we can evaluate the sought function

$$\left(\frac{40}{2/3}\right)^{2/3}\left(\frac{40}{1/3}\right)^{1/3}\left(\frac{1/2}{1/3}\right)\left(\frac{1/4}{1/4}\right)(2/3)^{2/3}=60$$

Thus, the minimum of the function $u(x_0)$ should be at least 60, and it is actually 60. However, it is not at all easy to prove that the value of $u(x_0)$ is its true minimum. This can only be done by use of a more general theory based on the fundamental theorem of criterion programming.

3.26. Primal and Dual Problems

Before we show that criterion programming can be used to solve optimization problems subject to inequality constraints, we shall formulate the primal and dual problems of criterion programming.

Primal problem. Find the minimum of the function

$$u(x), \qquad x=(x_1, \ldots, x_n)$$

subject to the constraints

$$x_1>0, \qquad x_2>0, \ldots, x_n>0 \tag{3.187}$$

$$u_1(x)\leqslant 1, \qquad u_2(x)\leqslant 1, \ldots, u_k(x)\leqslant 1 \tag{3.188}$$

It is also assumed that both the function to be minimized, $u(x)$, and the constraint $u_i(x)$ are generalized posynomials of the form of Eq. (3.155).

The exponents α_{ij} of the variables being optimized may be any real numbers, while the generalized constants A_i may only be positive. The posynomial to be minimized, $u(x)$, is called the *primal function*, and the variables x_j are called the *objective variables*. The constraints, Eqs. (3.187), are called *natural*, and the constraints, Eq. (3.188), are *forced constraints*. Together, the two are primal constraints.

For every primal problem in criterion programming there is another problem which is its dual.

Dual problem. Find the maximum value of a multiplicative function

$$D(\pi), \qquad \pi=(\pi_1, \ldots, \pi_m)$$

subject to the following constraints

$$\pi_1>0, \qquad \pi_2>0, \ldots, \pi_m>0 \tag{3.189}$$

$$\sum_{i\in J[0]} \pi_i=1 \tag{3.190}$$

$$\sum_{i\in J[t]} \alpha_{ij}\pi_i=0, \qquad j=\overline{|}\,{}^{n}_{1}, \quad t=\overline{|}\,{}^{k}_{1} \tag{3.191}$$

The multiplicative function may be written as

$$D(\pi)=\left[\prod_{i=1}^{m}\left(\frac{A_i}{\pi_i}\right)^{\pi_i}\right]\prod_{t=1}^{k}[\lambda_t(\pi)]^{\lambda_t(\pi)} \tag{3.192}$$

where

$$\lambda_t(\pi)=\sum_{i\in J[t]}\pi_i$$

The constants A_i are deemed positive and the coefficients α_{ij} are real numbers.

The function $D(\pi)$ is called a *dual function*, and the variables π_i are *dual variables*. The constraints, Eq. (3.189), are referred to as *positiveness conditions*, the constraints, Eq. (3.190), are *normality conditions*, and Eq. (3.191), the *orthogonality condition*. Together, they form dual constraints.

In criterion programming, it is assumed that a problem (primal or dual) is feasible if there exists at least one point (vector) satisfying all constraints. The primal problem A is said to be strongly feasible if there exists at least one vector x having positive constants and satisfying the condition

$$u_t(x)<1$$

A primal problem may be feasible without being strongly feasible, but every strongly feasible problem is feasible.

3.27. Fundamental Theorem of Criterion Programming

It can be proved that, subject to some constraints imposed on the choice of the variables π_i, the minimum of a primal function is the maximum of its dual function. This is proved by the fundamental theorem of criterion programming, which is stated as follows*.

Suppose that a primal problem is strongly feasible and that the primal function $u(x)$ takes on a minimum value at a point satisfying the primal constraints. Then:

(1) its dual problem is feasible and the dual function $D(\pi)$ is a maximum at a point satisfying the dual constraints;

(2) the maximum value of the dual function (subject to the constraints) is equal to the minimum value of the primal function $u(x)$ (also subject to the constraints);

(3) if x_0 is an optimal point in the primal problem, there exist nonnegative Lagrangian multipliers μ_{0t}, $t=\overline{1,k}$, such that the Lagrangian function

$$L(x,\mu)=u(x)+\sum_{t=1}^{k}\mu_t[u_t(x)-1] \tag{3.193}$$

has the property

$$L(x_0,\mu)\leqslant u(x_0)=L(x_0,\mu_0)\leqslant L(x,\mu_0) \tag{3.194}$$

* The proof of the fundamental theorem is given in [3].

for arbitrary $x_{j0}>0$ and arbitrary $\mu_t>0$. Also, there exists a maximizing vector π_0 for the dual problem with components

$$\pi_0=\begin{cases} A_i x_1^{\alpha_{i1}}, \ldots, x_n^{\alpha_{in}}, & i\in J[0] \\ \mu_t A_i x_1^{\alpha_{i1}}, \ldots, x_n^{\alpha_{in}}, & i\in J[t] \end{cases} \tag{3.195}$$

where $x=x_0$ and $\mu_t=\mu_{0t}$.

Besides,

$$\lambda_t(\pi_0)=\mu_{0t}/u(x_0), \qquad t=\begin{vmatrix} k \\ 1 \end{vmatrix} \tag{3.196}$$

(4) if π_0 is a maximizing point in the dual problem, then each maximizing point x_0 in the primal problem satisfies the following system of equations:

$$A_i x_{10}^{\alpha_{i1}}, \ldots, x_{n0}^{\alpha_{in}}=\begin{cases} \pi_{i0} D(\pi_0), & i\in J[0] \\ \pi_{i0}/\lambda_t \pi_0, & i\in J[t] \end{cases} \tag{3.197}$$

where t takes on any positive integer value for which $\lambda_t(\pi_0)>0$.

Relation (3.195) may be used to evaluate the maximizing vector π_0 from the known minimizing vector x_0 and the corresponding Lagrangian multipliers μ_{0t}. Conversely, relation (3.197) can be used to evaluate the minimizing vector x_0 from the known maximizing vector π_0 and the constants A_i. From analysis of Eq. (3.196) it is seen that the numbers $\lambda_t(\pi_0)$ accurate up to a constant factor are Lagrangian multipliers for the primal problem.

3.28. Procedure for Solving Constrained Problems

As an example, we shall solve the following problem. Suppose we want to minimize the posynomial

$$u(x)=40x_1x_2+20x_2x_3$$

subject to the constraints

$$u_1(x)=\frac{1}{5}x_1^{-1}x_2^{-1/2}+\frac{3}{5}x_2^{-1}x_3^{-2/3}\leqslant 1$$

It is to be noted that the problem involves four terms and three objective variables.

The dual function, Eq. (3.192), for this problem may be defined as

$$D(\pi_0)=\left(\frac{40}{\pi_{10}}\right)^{\pi_{10}}\left(\frac{20}{\pi_{20}}\right)^{\pi_{20}}\left(\frac{1/5}{\pi_{30}}\right)^{\pi_{30}}\left(\frac{3/5}{\pi_{40}}\right)^{\pi_{40}}(\pi_{30}+\pi_{40})^{\pi_{30}+\pi_{40}}$$

The normality condition, Eq. (3.190), for the primal similarity criteria may be written as

$$\pi_{10}+\pi_{20}=1$$

and the orthogonality condition, Eq. (3.191), for all similarity criteria as

$$\pi_{10}-\pi_{30}=0, \qquad \pi_{10}+\pi_{20}-\pi_{30}/2-\pi_{40}=0$$

$$\pi_{20}-2\pi_{40}/3=0$$

The linear system of four equations in four unknowns π_0 can be used to evaluate the similarity criteria:

$$\pi_{10}=1/2, \qquad \pi_{20}=1/2, \qquad \pi_{30}=1/2, \qquad \pi_{40}=3/4$$

Substituting these values in the dual function yields

$$D(\pi_0)=40$$

According to the fundamental theorem of criterion programming, the minimum value of the function $u(x_0)$ subject to constraints is equal to the computed value of the dual function:

$$u(x_0)=D(\pi_0)=u(x_{10}, x_{20}, x_{30})=40$$

Thus, we have evaluated the function $u(x_0)$ without having determined x_0.

To evaluate the vector x_0, we need three equations. According to the fourth point of the fundamental theorem [see Eq. (3.197)], the numbers π_{10} and π_{20} multiplied by $D(\pi_0)=u(x_0)$ yield the first and second terms of the objective function for an optimal alternative:

$$40x_{10}x_{20}=(1/2)\times 40, \qquad 20x_{20}x_{30}=(1/2)\times 40 \tag{3.198}$$

while the first and second terms in the constraint $u_i(x)$ respectively are

$$\pi_{30}/(\pi_{30}+\pi_{40})=2/5$$
$$\pi_{40}/(\pi_{30}+\pi_{40})=3/5$$

So, in accordance with Eq. (3.197) we get

$$\begin{aligned} \frac{1}{5}x_{10}^{-1}x_{20}^{-1/2}&=2/5 \\ \frac{3}{5}x_{20}^{-1}x_{30}^{-2/3}&=3/5 \end{aligned} \tag{3.199}$$

Take logarithms of Eqs. (3.198) and (3.199):

$$\begin{aligned} \ln x_{10}+\ln x_{20} &= -\ln 2 \\ \ln x_{20}+\ln x_{30} &= 0 \\ -\ln x_{10}-\frac{1}{2}\ln x_{20} &= \ln 2 \\ -\ln x_{20}-\frac{2}{3}\ln x_{30} &= 0 \end{aligned}$$

In this linear system of equations written for $\ln x_{j0}$, one equation is linearly dependent. Solving this system gives

$$\ln x_{10}=-\ln 2, \qquad \ln x_{20}=0, \qquad \ln x_{30}=0$$

Therefore, $$x_{10}=1/2, \qquad x_{20}=1, \qquad x_{30}=1$$

Thus, the variables of the optimal alternative are:

$$x_{10}=1/2, \qquad x_{20}=1, \qquad x_{30}=1, \qquad u(x_0)=40$$

3.29. Modified Lagrangian Function

In Sec. 3.28 we have shown that criterion analysis can be used to determine optimal alternatives, assuming that functional inequality constraints are imposed on the objective variables.

However, investigation has shown that criterion programming is only applicable to technico-economic models for which the so-called modified Lagrangian function (to be defined later) has canonical form. Now we shall explain what this means.

We set out by assuming that the objective function (one to be minimized) is a posynomial of the form

$$u(x)=\sum_{i=1}^{m} A_i \prod_{j=1}^{m} x_j^{\alpha_{ij}} \tag{3.200}$$

and changes in the objective variables are subject to k functional constraints each of which is likewise a posynomial:

$$G_t(x)=\sum_{i\in J[t]} \beta_i \prod_{j=1}^{n} x_j^{\alpha_{ij}}=1, \qquad t=\overline{|1}^{k} \tag{3.201}$$

Since these are equality constraints, the optimal alternative can be found by the method of undetermined multipliers called Lagrangian multipliers. For this purpose, we write the Lagrangian function

$$L(x)=\left(\sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_j^{\alpha_{ij}}\right)+\sum_{t=1}^{k} \mu_t \left(\sum_{i\in J[t]} \beta_i \prod_{j=1}^{n} x_j^{\alpha_{ij}}-1\right) \tag{3.202}$$

where μ_t are the undetermined, or Lagrangian, multipliers.

Owing to the introduction of as many multipliers μ_t as there are constraints, the Lagrangian function in the form given above makes it possible to reduce the maximization of the function (3.200) subject to the constraints (3.201) to the minimization of the function (3.202) free from any constraints.

The optimal values of the objective variables thus found must satisfy the constraints (3.201), or

$$\sum_{i\in J[t]} \beta_i \prod_{j=1}^{n} x_{j0}^{\alpha_{ij}}-1=0, \qquad t=\overline{|1}^{k}$$

So when the variables have optimal values the Lagrangian function, (3.202), yields the same value as the original objective function

$$L(x_0)=u(x_0)$$

To facilitate further computations, the Lagrangian function, (3.202), may conveniently be re-written as

$$L(x)+\sum_{t=1}^{k} \mu_t=\sum_{i=1}^{m} A_i \prod_{j=1}^{n} x_j^{\alpha_{ij}}+\sum_{t=1}^{k} \sum_{i\in J[t]} \mu_t \beta_i \prod_{j=1}^{n} x_j^{\alpha_{ij}} \tag{3.203}$$

The above transformation is analogous to transposing the constant cost items to the left-hand side of the equation, as was done in Eq. (3.30). The expression thus derived will be called the *modified Lagrangian function*

$$\mu(x)=L(x)+\sum_{t=1}^{k}\mu_t \tag{3.204}$$

The function, Eq. (3.204), has as many terms as there are terms in the objective function $u(x)$ plus the total number of terms in the functional constraints, Eq. (3.201). Let this number be γ. Then the modified Lagrangian function, Eq. (3.204), may be written as

$$\mu(x)=\sum_{i=1}^{\gamma}B_i\prod_{j=1}^{n}x_j^{\alpha_{ij}} \tag{3.205}$$

The above function differs from the discounted cost equation, Eq. (3.16), only in some notation. Therefore, if this function is canonical, the similarity criteria of optimal alternatives will be independent of the numerical values of the design data, or constants, B_i, and all of the techniques of criterion analysis applicable to canonical equations may be used as well.

Thus, criterion analysis may be used under conditions of constraints imposed on the decision or objective variables, provided the modified Lagrangian function has canonical form, too.

The difference in the analysis of canonical discounted cost equations and of canonical modified Lagrangian functions consists in that instead of an economical alternative one uses an optimal alternative as the reference. As a result, it is possible to use methods based on the theory of similarity not only to find the economically most advantageous alternative, but also to establish the contributions of the various cost items and to test the design for sensitivity and economic stability subject to technical constraints.

Sometimes, problems to which the modified Lagrangian function applies are called zero-degree-of-difficulty problems. Generally, the degree of difficulty, ϱ, of a problem in criterion analysis is defined as

$$\varrho=\gamma-n-1$$

At present, detailed techniques have been developed for solving only zero-degree-of-difficulty cases, although there exists in criterion analysis an acute need for techniques that could handle problems of a positive degree of difficulty.

REFERENCES TO CHAPTER 3

1. Wentzel E. S. *Operations Research* (Issledovanie operatsy). Sovetskoye Radio, 1972.
2. Kuzin L. T. *Fundamentals of Cybernetics* (Osnovy kibernetiki). "Energiya", 1973.
3. Duffin R. L., Peterson E. L., Zener K. *Geometric Programming, Theory and Application.* John Wiley & Sons, Inc. New-York, London, Sidney, 1967.
4. Venikov V. A. *Theory of Similarity and Modelling with Special Reference to Power System Problems* (Teoriya podobiya i modelirovaniya primenitelno k zadacham elektroenergetiki). "Vysshaya Shkola", 1966.

CHAPTER 4

INFORMATIONAL ANALYSIS OF CONTROL AND MANAGEMENT IN ELECTRIC POWER SYSTEMS

4.1. Information and Degree of Organization Defined

Information reflects an isolated essence or characteristic of a concrete thing or an event. The word information is usually associated with a message, a thing told, or items of knowledge.

The concept of information is related to certain models of real things, models which reflect their character to the extent necessary for practical purposes. This approach to information agrees with the philosophic concept of reflection according to which information reflects both the real properties of objects, independent of a subject, in animate and inanimate nature, and the properties of cognition and thinking.

To sum up, information should be construed as the essential and most representative characteristics or the essence revealed in a well-defined manner, of things and processes in the material world, and not the things and processes themselves. In other words, what one has in mind when one speaks of information are not things or processes as such, but their representations in the form of numbers, equations, descriptions, symbols, images, or other abstract characteristics.

The transmission and processing of information is involved in the activity of any automatic device, living organisms, man, economic and social systems, and life itself. Science, too, may be looked upon as a man-made complex and developing information system.

Information may be classed by fields of knowledge (biologic, technical, economic, etc.), by the physiological mechanism involved in its perception (visual, aural, tactile, etc.), by the manner in which it appears in control processes (sensing, recognition, prediction, decision-making, etc.), or by the metrical properties utilized (parametric, topological or abstract).

Parametric information includes sets of values that measured variables can take, the results of determinations made in the course of investigations, analysis, or accounting. Parametric information is widely used in science and technology to present the results of measurement.

Topological information covers geometric patterns, maps, pictures and three-dimensional representations. This form of information is convenient where patterns and situations are to be suitably recognized.

Abstract information takes the form of mathematical relations, generalized concepts and ideas. This type of information is usually resorted to

where abstraction, generalization and symbolism are needed in theoretical studies.

The following phases can be singled out in the circulation of information: generation, transmission, reception, processing, presentation and utilization. Since information is carried by signals, an information circulation cycle may be treated as a circulation and processing cycle of information-carrying signals.

Any object as a source (generator) of information is inexhaustible. Some of the information it generates is converted by sensing elements into signals widely varying in physical nature: mechanical, optical, electrical or electromagnetic.

As an example, consider the case where signals from a source, *S*, are conveyed by a radio link, *RL* (Fig. 4.1). The signals coming from transducers, *Ts*, are coded by an encoder *EC* into electrical signals which are superimposed on the carrier emitted by the transmitter, *Tr*, to change, or modulate, its amplitude, frequency or phase. The receiver, *R*, accepts both the useful signal and noise. The signal is then separated from noise and decoded (normalized, quantized, sampled, etc.) by a decoder *DC*, for further processing, *Pr*.

Signal processing is done by devices or computers which convert the incoming quantities and functions from analog to digital or from digital to analog form. An intermediate stage in processing is the storage of information in storage or memory devices which may be magnetic discs, ferrite cores, magnetic drums, magnetic tape, various cards, relays, electrical, ferro-electric, electroluminescent or cryiogenic elements, delay lines, diode and plug matrices.

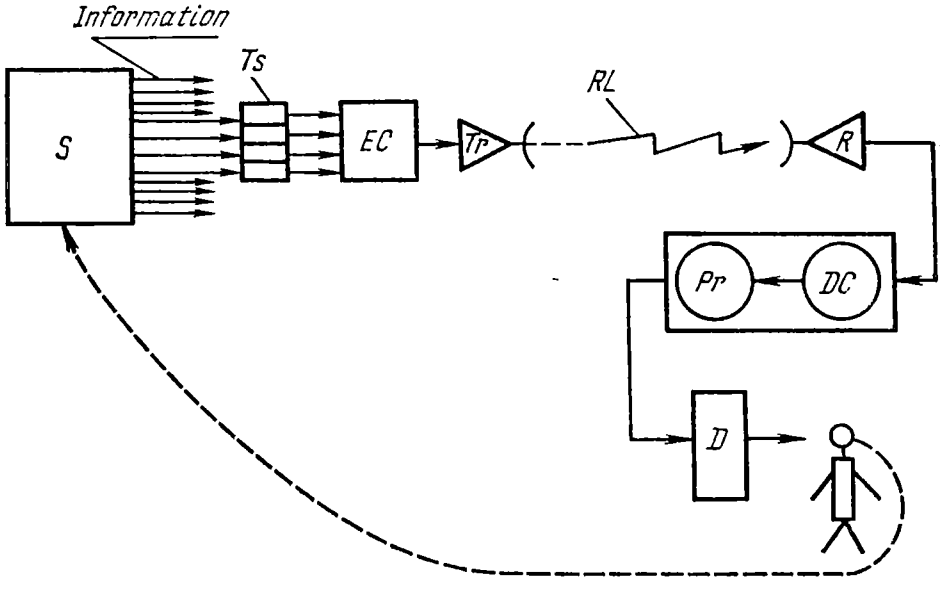


Fig. 4.1

A very important element in a man-machine information processing system is a display, *D*. If there is a human operator in a control loop (an interactive system), the symbols chosen to represent the qualitative and quantitative aspects of processed information must be displayed in a most optimal manner so that the operator can make proper decisions, especially under time limitations. This purpose can be served by a variety of devices: optical, acoustical, alphameric, deflection and graphic indicators; digital and graphic recorders; CRT oscilloscopes, video displays; two- and three-dimensional mimic panels; light panels and mock-ups with built-in indicating elements.

Information can be utilized in one of two ways: by execution or implementation, where the information received is transformed into material and power actions; and by extraction of contents, which depends on the qualification of the user and his ability to utilize information in achieving his goals.

If the system under study is complex, the need may arise for simplifying it. In such a case, information stands out as a methodological basis for simplification. In the words of W. R. Ashby, one of the founders of cybernetics, information theory itself may be looked upon as a form of simplification, because instead of tracing every individual cause to its individual effect, which is a classical element of scientific inquiry, it lumps together all causes and all effects and only relates the two totals.

Much as the concept of energy has led to a unified and progressive point of view on physical processes in nature and technology, so, by analogy, the concept of information may lead to a unified approach to the analysis of qualitative and quantitative aspects of the processes involved in thinking, signal perception, and control.

While the laws governing the transformations of matter and energy are well known today, those related to the generation and transformation of information are little known, and we are still lacking a universal mathematics for their description. In fact, we are still lacking a generally approved and single definition of information, although its interpretations are many. In our discussion, we shall use an interpretation which reflects the duality of its properties and relates it to another important concept, degree of organization. Thus, information is an attribute of matter, acting as a characteristic of the degree of organization of matter, on one hand, and as a means of achieving its organization, on the other. As a means of changing the degree of organization of matter, information is of potential nature if there are no means to implement it in the form of material or power actions on matter, or both.

Degree of organization is one of the fundamental properties of all material systems, from the atom to any large complex. This is only too natural, because any real system has structure. It is to be noted, however, that not just any structure is a system (a system differs from a structure in that it is isolated from the surroundings, and the criterion is the difference in degree of organization between the system and the surroundings). Degree

of organization can characterize the structure, set-up and properties of a system. An important form of degree of organization is goal-oriented organization, that is, organization with respect to a particular goal. This can only occur in biological, social and man-made systems, that is, in systems which depend for their functioning on control processes. Goal-oriented organization is defined both by the quantity of useful information contained in the behaviour program, and by the capacity of organizational means. Here, *organizational means* refer to means of implementation (human or machine) which execute the behaviour program as activities directed towards the goal of control, and the *capacity of organizational means* refers to their potential ability to act upon the system.

As an intuitive concept, degree of organization has gained wide-spread currency. However, we are still lacking an established formalized definition of degree of organization. Existing definitions are either philosophical or based on the assumption that degree of organization can be determined in terms of the quantity of statistical or combinatorial information. In such an approach, degree of organization reflects only a special case out of a multitude of forms in systems.

Since a system and its constituent elements must be viewed from three stand-points — structure, set-up and properties — degree of organization too will manifest itself in three ways — in structure, set-up and properties. Also, within these basic gradations, degree of organization may be temporal, spatial, algorithmic or statistical in form. Furthermore, it may be relative or absolute, external or internal, goal-oriented or not.

As will be shown presently, the degree of organization is a function of the simpler characteristics of a system — complexity and orderliness.

4.2. Basic Qualitative Aspects of Information in Control and Management Systems

In the operation of a system, information may be present in various qualitative forms. Usually, only its passive form is implied, that is, the reflection of the properties of a system. As already noted, however, information may be active by causing changes in the structure, set-up or properties of a system.

In the operation of systems, we shall recognize the following distinctive forms of information: input, I_{in}, processing; I_p, processed and control(ling), I_c. In cybernetic and biological systems processed information should further be subdivided into sensing information, I_s, recognition information, I_r, decision-making information, I_{dm}, prediction information, I_{pred}, and execution information (implementation by organizational means, OM, that is, I_e). These forms of information are shown in the block diagram of a control system in Fig. 4.2.

In the general case, input information covers all objective information, I_{in}^{o}, about the structure, set-up and properties of the controlled plant or process, CP, and also about the environment, I_{in}^{e}. A special case of input

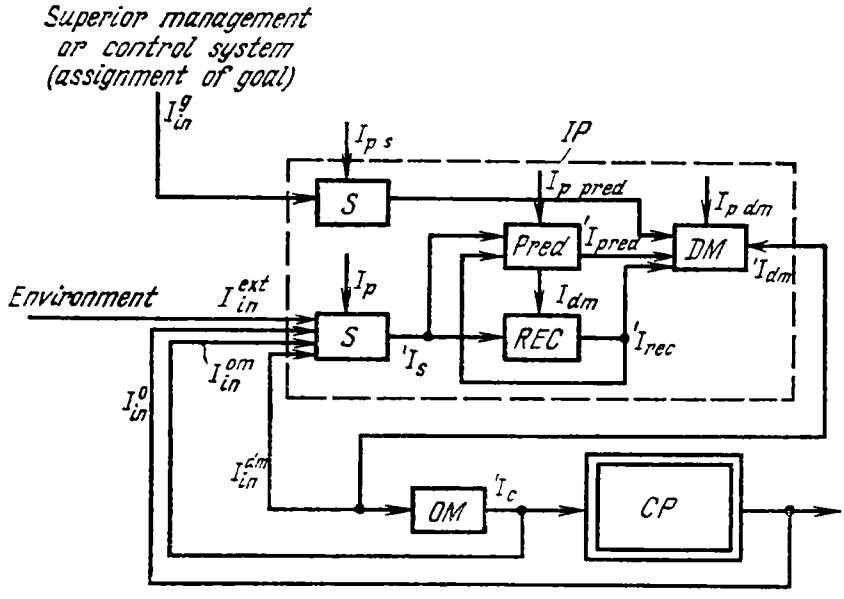


Fig. 4.2

information is that about the goal of the system's functioning, I^g_{in}, which may in turn be objective or subjective, and that about the decisions made, I^{dm}_{in} and about the action of organizational means I^{om}_{in}.

Processing information, I_{ps}, I_{pr}, $P_{p\,pred}$ and $I_{p\,dm}$, characterizes the structure, set-up and properties of the information processor, *IP*, which, in the general case, is a control centre employing human staff and machines in accordance with the goals sought. In a special case and for purposes of information analysis, an information processor may be represented simply by the algorithms used to process the incoming information. Processing information has to do with the form of processing and the goal of system operation, because it is a means of extracting useful information (about form of processing and utilization of information) from input information.

Processed information is in fact input information reduced in a manner decided by the form of processing accomplished by the information processor. Processed information may be objective or subjective, according to the type of processor used. It is always a function of the form (task) of processing. It is also a function of the final goal, this relation being stated in terms of indicators of the value (utility) that information may have in achieving goals specified objectively or subjectively.

Control(ling) information covers all information which is utilized by organizational means as a guide (cause) for their action to change the state of the controlled plant or process, in the general case, its structure, set-up or properties. Processed information can be transformed into control(ling) information by looping it to the controlled plant or process via organizational means. As already noted, in engineering systems the organizational means are actuators or final control elements. Controlling information is goal-oriented, that is, related to the goal that the system seeks to attain

by its functioning. It may be said that controlling information is a means of optimizing system operation with regard to the goals sought.

As already stated, the overall process of control or management may be visualized as involving sensing, recognition, prediction, decision-making and execution, and each of these stages as utilizing its own form of information. Accordingly, sensing information may be thought of as a representation of the useful aspects of input information reflected in the elements of the information processor (the sensing unit); recognition information as a representation of the situations defined by the input information on the finite set of patterns specified by the elements of the information processor (the recognition unit); prediction information as a characteristic of the system state or behaviour which has a certain probability of taking place with a specified anticipation time t_{pred} (this information is mapped on a finite set of elements which make up the prediction unit in the information processor); decision-making information as a mapping of patterns and goals (both current and predicted) onto the finite set of decisions specified by the elements of the information processor (the decision-making unit). Control-(ling) information is a special form of information, because it is a cause for action by organizational means and, through them, for changes in the state or behaviour of the controlled system.

Information analysis of necessity raises the question of the objectivity and subjectivity of information and of the relations between the object and the subject.

In philosophy, the term subjective defines a judgement passed on an object by an individual and generally not true to reality; that is, it is stressed that the judgement is in error and has its source in the mind of a subject. Here "subject" defines a thinking person in the process of learning the laws that govern the objective world and trying to change it on the basis of what he learns. Lately the concept of machine intellect has been gaining popularity and it has become possible to model at least some psychic types of the nervous activity and the mechanism of emotions. Because of this, it is no longer justified to set man in opposition to machine as regards information. Instead, the concept of subject is increasingly more often treated from a cybernetic point of view and it is assumed that any informational structure able to perceive information from an object, classify situations in it, predict its behaviour, and make decisions leading the object to a wanted goal may be counted as a subject. Thus, a subject should not necessarily be a human being, and the very concept becomes a relative one to some extent.

As an example, consider a case where the object is a patient and the subject is a diagnostic and cure-selecting system*. If this is an automatic system, the subject will be a machine. Practically, such a system will always be supervised by a human physician to whom both the patient and the machine are objects (Fig. 4.3). It appears practical to class objects and

* By the same token, this may be a system to diagnose faults in an electric system.

subjects as those of first and second order. Symbolically, the relations between the patient, *Pt*, the machine, *M*, and the physician, *Pn*, may be defined as

$$O'(Pt) \rightarrow S'(M); \qquad O''(Pt, M) \rightarrow S''(Pn)$$

where

$$O''(Pt, M) = O'(Pt) + S'(M)$$

In the general case, information between the machine, *M*, and man, *H*, can be transferred in any one of four principal ways:

$$O(H) \rightarrow S(M); \quad O(M) \rightarrow S(M); \quad O(H) \rightarrow S(H); \quad O(M) \rightarrow S(H)$$

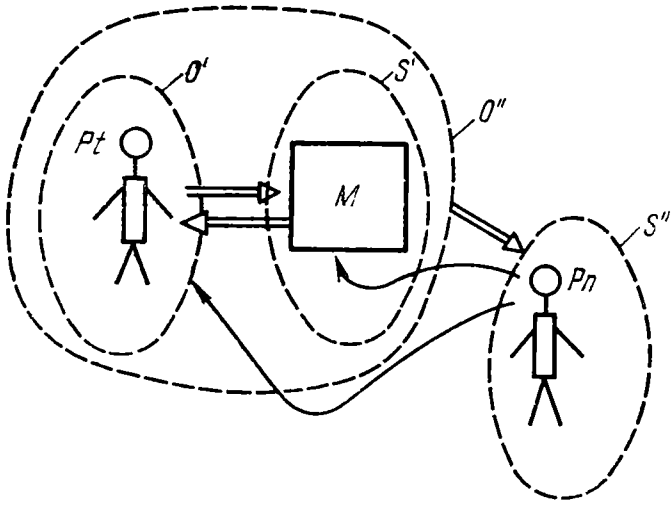


Fig. 4.3

Each subject will show an individuality of his own, only because systems with a complex structure will always be different to a varying degree. Also, the decisions made by a subject will always be other than optimal, because an object is only partly cognizable, partly observable and partly controllable for a subject. However, the degree of subjectivity can be different, and it makes sense to class information or a goal as subjective only when this degree is considerable (on goal subjectivity, see Sec. 4.8).

4.3. Brief Outline of the Statistical Theory of Information

In the probabilistic approach, information is regarded as a message about the outcome of random events. The greater the uncertainty of an event, the greater the freedom of choice in forming a message and the greater the amount of information it can carry. Messages about frequent events for which the probability of occurrence tends to unity carry little information, and so do messages about the non-occurrence of an event which was hardly likely to happen anyhow, that is, one for which the a priori probability is close to zero or, which is the same, the probability of an anti-event

is close to unity. Thus, an event and its dual anti-event make up a single binary event. For example, that an electric power transmission line is damaged is an event and that it is not damaged is an anti-event. Such an event may occur with a probability $p=0$ to 1 or, accordingly, not occur with a probability $q=1-p$. Events may also be regarded as n probable outcomes of an experiment with these outcomes making up an *ensemble* or a complete system of statistically independent events for which the sum of probabilities is

$$\sum_{i=1}^{n} p_i = 1$$

It is to be noted that one and only one event out of the complete set or ensemble must occur at each trial.

If the probabilities of events are held constant, the uncertainty of an ensemble is qualitatively expressed as its entropy. According to Shannon, the entropy of a complete set of events is the average function of the probabilities of each likely outcome.

Let N trials result in k distinct outcomes. The amount of information in each ith outcome will depend on the a priori probability of that event and can be defined in binary units (bits or binits) as

$$I_i = \log_2 (1/p_i) = -\log_2 p_i \tag{4.1}$$

If $N \gg n$, then each ith outcome should occur n_i times in the course of all trials. Then the average information yielded by one trial will be

$$I_{av} = \frac{n_1 I_1 + n_2 I_2 + \ldots + n_k I_k}{N}$$

$$= (n_1/N)(-\log_2 p_1) + (n_2/N)(-\log_2 p_2 + \ldots + (n_k/N)(-\log_2 p_k) \tag{4.2}$$

Since n_i/N is the frequency at which the individual outcomes occur, replacing n_i/N in Eq. (4.2) with the probability p_i yields

$$H = I_{av} = -\sum_{i=1}^{k} p_i \log_2 p_i \tag{4.3}$$

Entropy is always a nonnegative quantity and is equal to zero when the probability of one event is unity and those of the remaining events are zero, that is, when everything is known in advance and the outcomes of trials can yield no new information. Entropy is a maximum when all probabilities are the same:

$$p_1 = p_2 = \ldots = p_k = 1/k$$

In such a case, the value of entropy is the same as the additive logarithmic measure of information proposed by R. V. L. Hartley

$$H = \log k \tag{4.4}$$

In control application, it is often necessary to measure the mutual information relating, say, input signals $x_1, x_2, \ldots, x_m$ of probabilities $p(x_1)$,

$p(x_2), \ldots, p(x_m)$, to the values $y_1, y_2, \ldots, y_m$ of some controlled output variable, of probabilities $p(y_1), p(y_2), \ldots, p(y_m)$.

If the input and output of a system are correlated, the statistical relation between them can be expressed in terms of conditional probabilities. For example, the probability of a specified signal x_i taking a value y will be defined by the conditional probabilities $p(y_1/x_i), p(y_2/x_i), \ldots, p(y_j/x_i), \ldots, p(y_m/x_i)$, where

$$p(y_j/x_i)=p(x_i, y_j)/p(x_i)$$

The partial conditional entropy (per signal x_i) is

$$H(Y/x_i)=-\sum_{j=1}^{m} p(y_j/x_i) \log (y_j/x_i) \tag{4.5}$$

Then the average entropy for all input signals x is

$$H(Y/X)=-\sum_{i=1}^{m} p(x_i)H(Y/x_i) \tag{4.6}$$

or, written out in full,

$$H(Y/X)=-\sum_{j=1}^{m}\sum_{i=1}^{m} p(x_i, y_i) \log p(y_j/x_i) \tag{4.7}$$

Obviously, with strong correlation, the signal x_i will correspond to a certain definite value of the output variable, say y. Then

$$p(y_j/x_i)=1, \qquad p(y_1/x_i)=p(y_2/x_i)=\ldots=p(y_m/x_i)=0$$

In such a case, $H(Y/X)=H(X/Y)$, and the amount of mutual information is equal to the unconditional entropy

$$I(X \cdot Y)=H(X \cdot Y)=H(X)=H(Y)$$

where the unconditional entropy of the ensemble X is defined as

$$H(X)=-\sum_{i=1}^{m} p(x_i) \log p(x_i) \tag{4.8}$$

and the unconditional entropy of the ensemble Y is

$$H(Y)=-\sum_{j=1}^{m} p(y_j) \log p(y_j) \tag{4.9}$$

For a complex event, a measure of uncertainty is given by the joint entropy

$$H(X, Y)=H(X)+H(Y)-H(X \cdot Y) \tag{4.10}$$

In turn, the mutual entropy is

$$H(X \cdot Y)=H(X)-H(X/Y)=H(Y)-H(Y/X) \tag{4.11}$$

If X and Y are not correlated, then obviously

$$H(X, Y)=H(X)+H(Y)$$

The amount of information is equal to entropy only when the uncertainty of a system (situation) is fully removed by receipt of a message. Otherwise, the amount of information in a message will be equal to the difference between the original entropy, that is, before receipt of the message, H_b, and the final entropy left after receipt of the message, H_a:

$$I=H_b-H_a \tag{4.12}$$

Transmission of information involves matching the source to the channel, selecting the best code, and estimating the necessary accuracy for the display. In these matters, use is made of the following expression which defines the relative redundancy of the source:

$$D=1-H/H_{\max} \tag{4.13}$$

This can be illustrated by the following example. A CRT display is capable of presenting the variable of interest in 1024 vertical gradations (quanta). However, the actual range of the variable encompasses 128 quanta. Therefore we may write for our example:

$$H_{\max}=\log_2 1024=10 \text{ bits}$$

and

$$H=\log_2 128=7 \text{ bits}$$

The redundancy of the display is thus:

$$D=1-7/10=0.3$$

Controlled variables are always measured with some approximation decided by the accuracy of the respective sensor or transducer. Therefore, the accuracy of the measuring instrument should be matched to the specified accuracy of measurement.

As an example, let the specified accuracy of measurement be $\pm\sigma$, and the resolution of the measuring instrument be

$$\Delta y=2\sigma$$

For the total range of measurements, $y_{\max}-y_{\min}$, the discrete values will be equal to a finite number

$$n=(y_{\max}-y_{\min})/\Delta y+1 \tag{4.14}$$

If all discrete values y have the same probability, the information content of each instrument indication will be

$$I=\log_2 n \tag{4.15}$$

For a digital readout instrument (for simplicity, one using binary coding), the number of digits N necessary to give the desired Δy must be such that

$$N \geqslant I$$

The maximum rate at which the instrument can generate information (in bits per second) is given by

$$C=\log_2 n/T \tag{4.16}$$

where T is the time constant of the instrument.

When Δy is specified in advance, the least time interval permissible in measuring y is

$$\Delta t=\Delta y/(dy/dt)_{\max} \tag{4.17}$$

where $(dy/dt)_{\max}$ is the maximum rate of measurement.

The least rate (bits per second) at which measurements can be made in such a case is

$$f=I/\Delta t \tag{4.18}$$

4.4. Hierarchy of Tasks, Goals, Algorithms and Control Centres

If a large system is functioning, this above all means that its control (information) centres are handling a multitude of control tasks in order to achieve specified goals and sub-goals with a certain degree of accuracy.

A few remarks are in order as regards the concepts of task, goal and subgoal and their inter-relations.

In the most general form, a task may be visualized as an activity carried on to overcome some conflicting requirements. Some tasks in a system are achieved during the formative stage of system control or in the course of its developments; they are handled by system engineers or operations research (OR) people. Among other things, they investigate the system in question and formulate control algorithms and goals for operators and automatic systems. Other tasks are handled by operators and automatic systems in the course of automatic control.

On the basis of conflicting requirements involved, all tasks may qualitatively be classed into five basic types (the names used below are arbitrary):

(1) synthesis, where design data and an idealized (overall) goal are specified, and it is desired, subject to time and resource constraints imposed on the operation, to find an algorithm which would be a solution of the problem;

(2) primal problems, where design data, an algorithm and a practically achievable goal are specified in advance, and the solution of the problem subject to time and resource constraints imposed on the control itself is a result leading to the control goal (the goal can be achieved only if the specified design data and algorithm are optimal);

(3) dual problems, where the goal and algorithm are known in advance, and it is desired, subject to time and resource constraints imposed on the operation, to define design data which would lead to the goal (the design data may be subgoals, control actions, etc.);

(4) analysis, where the solution of one of the above problems is known, and it is desired subject to time and resource constraints imposed on the

operation, to compare the solution with a reference one; if the difference exceeds a preset limit, methods should be found to reduce it;

(5) realization problems, where design data, an algorithm and an idealized goal are known in advance, and it is desired to simplify the algorithm and formulate a practical goal by discarding insignificant factors, introducing indirect criteria, etc. This is done so that the primal problem can be solved directly under conditions of control, where more rigorous time and resource constraints are effective and there is a smaller number of observables.

The OR (operations research) team tackles tasks 1, 3 and 5 in developing a control system, and task 4 in optimizing such a system. Task 2 is the responsibility of the operator and automatic systems.

We have taken up the tasks arising in the functioning of a system. About the same approach can be taken in the case of system expansion, but the design data, algorithm and goals will be different in content, and the operator gives way to people responsible for development, engineering, installation, construction, and other work associated with an expanding system.

We shall demonstrate the relations that exist between an OR man and an automatic system operator by reference to the diagram of Fig. 4.4.

In accordance with the established hierarchy of tasks, the OR man is assigned a job — a task, an idealized goal, the constraints effective during the operation, and resources. By carrying out the operation, he is expected, firstly, to obtain an estimate of the internal and external conditions, compare alternative algorithms for achieving the assigned goal and select the

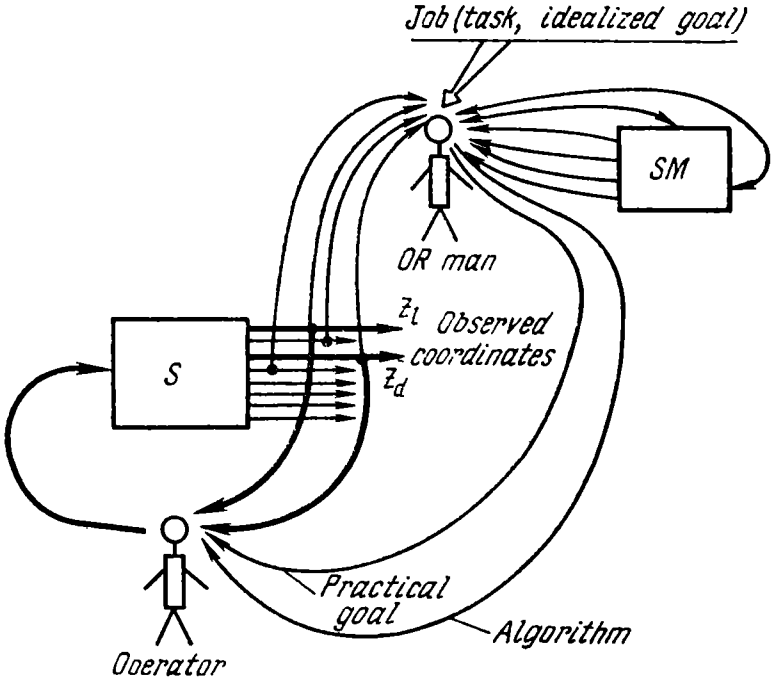


Fig. 4.4

optimal one; secondly, to formulate a practical goal and select the best algorithm for the operator, with due account of his capabilities and the constraints imposed on control.

As an example, the task and an idealized goal may be stated as follows: find a complex control algorithm which, in all the circumstances likely to arise in an electric power system due to changes in the basic variables $x_1, x_2, \ldots, x_m$ and the specified maximum disturbances $F_{1\max}, F_{2\max}, \ldots, F_{k\max}$, would enable the system to function so as to satisfy the quality criteria $W_1, W_2, \ldots, W_n$. Here, the idealized goal is a domain within the space of criteria $W_1, W_2, \ldots, W_n$.

In tackling his task, the OR man, in addition to the observable coordinates of the system, may use additional information such as that derived from, say, analysis of statistical data or, as is shown in Fig. 4.4, from studies on the system model, *SM*. The resources available to the OR man are not limited by any apparatus allocated, as is the case with the operator; if necessary, he may use a computer. Also, the time allotted to the operation is considerably greater than that during which the operator is to make a decision.

Bearing in mind the constraints to which the operator will be subjected and relying on the control algorithm already found and the idealized goal, the OR man establishes the best algorithm and formulates a practical goal for the system operator. In our example, a practical goal may be visualized as a domain within the space of observable coordinates Z_l, Z_d. Here, the practical goal appears as a mapping of an idealized goal from space W into space Z. This goal is a kind of target the operator is expected to hit with a representative point which is a solution to the periodically repeated problem of operational control.

Thus, a particular goal is achieved by solving a particular problem of system operation. In this context, the design conditions may be interpreted as subgoals subordinate to the ultimate goal. They may be specified in mathematical, logical or any other form. Satisfying the design conditions means that the ultimate goal can be achieved. In some cases, the ultimate goal would be achieved automatically as a result of meeting its subgoals, while in other cases this would require some additional effort (in terms of information, materials or energy).

Tasks, goals and subgoals along with an algorithm are the elements that define a control strategy for control centres. In terms of content, each of these elements may be stated as answers to some questions: a task — what should be done? a goal — what results are to be expected? subgoals — under what conditions? an algorithm — how should everything be done?

In a large system, tasks, goals, subgoals and algorithms on the one hand, and the control centres on the other are themselves complex entities which act as specific systems, for example, as a system of tasks, or as a system of goals. These systems are made up of hierarchically related subsystems (elements). The decomposition of a system into elements may proceed in any one of several ways, each leading to a different number of subsystems.

A system decomposed into subsystems which are then arranged in an appropriate hierarchy produces a system tree which has a definite number of levels or ranks (Fig. 4.5).

The operation of a large system can be represented as a tree of tasks which should be handled at each level of system control. Proceeding from the requirements for the quality of control, a control goal tree is set up. It is important to note that one task may involve several goals each of which should be achieved in the course of control.

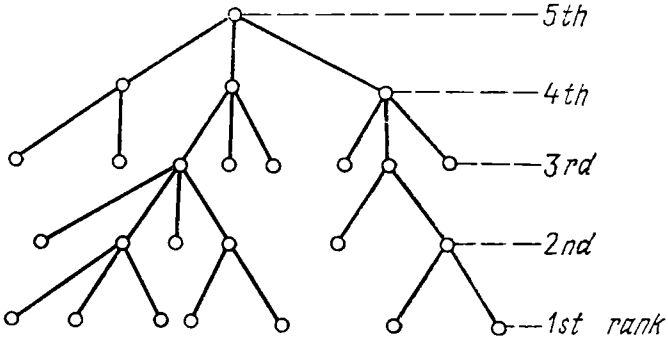


Fig. 4.5

If no quantitative conditions are formulated, the achievement of the task is the goal; if the results of realization are not quantifiable, the task and the goal merge together, as it were.

Qualitatively, goals may be classed into *mix* (set-up) *goals*, *property goals*, and *structure goals*. In the first case, the goal is to bring out the mix of the elements making up the system, or to identify the elements decisive to the behaviour of the system. In the second case, the goal is to define the structure of the relations between the system elements or the actions taken. In the third case, the goal is to establish various desirable characteristics, temporal, spatial, etc., which reflect the properties of the system or the actions to be taken.

The operation of a system as an entity can be defined in terms of one or several global (principal) goals. The global goals and the subgoals subordinate to them make up a hierarchical tree. If a global goal be visualized as the trunk of the tree, the subgoals will be its branches.

If a system has several global goals, no single global goal can be formulated within the respective goal hierarchy. However, this does not preclude the existence of another, higher system having a higher goal to which the global goals under consideration are subgoals. It is not unlikely that the global goals in the system under consideration may be independent or even conflicting.

For each hierarchy of tasks and of goals there should be a corresponding hierarchy of algorithms stated as mathematical, logical or verbal codes of rules each oriented towards the attainment of a goal (or goals) of its own.

In this context, 'of its own' is to be taken as requiring that each algorithm should provide for the attainment of one or several associated goals with some specified precision.

All of this hierarchy of algorithms is realized by various automatic controllers, computers and human operators concentrated at control centres on different hierarchical levels. Thus, the hierarchy of control centres materializes by its operation the first three hierarchies — those of tasks, goals and algorithms.

The attainment of global goals is secured by the higher levels of the control hierarchy, and that of the lower goals by the control centres at the lower levels. Each control centre within a control hierarchy acts so as to attain one or several goals and, in the final analysis, the global goal.

Qualitatively, the performance of a control centre may be assessed in terms of a function of the error between the goal and the representative point in the goal coordinate space. In Sec. 4.8, this function will be defined as the goal-related organization deficiency, or disorganization.

If we determine this disorganization for each goal, we can construct a goal-related disorganization hierarchy from which it is possible to identify, say, bottlenecks in the operation of the system.

The relations between the five hierarchies in a large system are shown in the diagram of Fig. 4.6. The hierarchy of goal-related disorganizations may be used as a kind of input information in optimizing the entire control system.

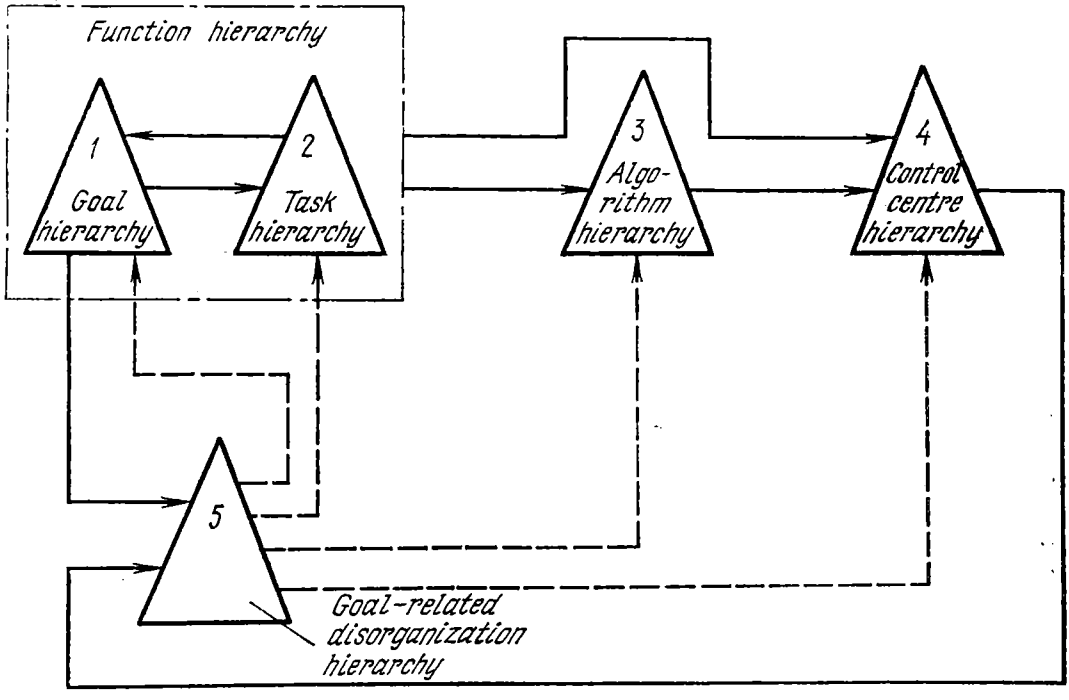


Fig. 4.6

4.5. Informational Description of Large System Operation

No single model is capable of representing a large system; to do so, a whole hierarchy of models is needed (Fig. 4.7), differing in the degree of detail involved. Four basic levels of detail may be singled out: the system-wide level, the subsystem level (there may be several such levels in the general case), the sectional level, and the element level. In the last case, the elements may be people, machines or algorithms.

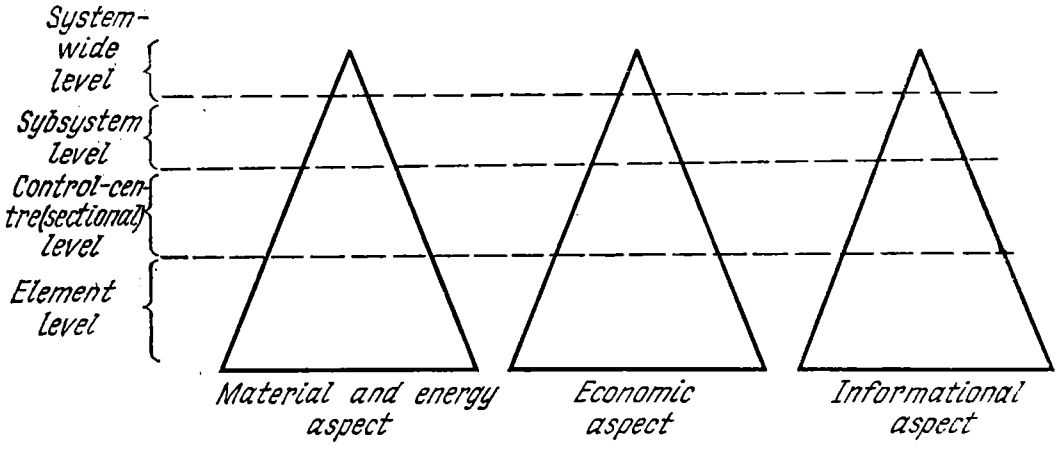


Fig. 4.7

Models can be classed not only vertically, but also horizontally into aspects of the events under study. In large man-made systems, three basic aspects can be identified: the material and energy aspect, the economic aspect, and the informational aspect.

The material and energy aspect covers various forms of material and energy flows and connections in the system of interest, and also the laws governing the transformations of energy and materials in the course of system operation and the forms that these transformations may take. This aspect is directly related to the technological processes occurring within the system and the attending events.

The economic aspect in the analysis of a large system is concerned with the effectiveness of system operation and expansion.

In the informational aspect, a system is examined from the standpoint of control of existing material, energy and economic processes. More specifically, the analyst is concerned with the relations existing between the various information centres and information processes within the centres. These centres may be controlled or controlling, depending on the hierarchical level involved.

These aspects are closely interrelated, because they are the facets of the same complex process — the activity of a system. It is only too natural, therefore, that no precise delineation is possible between them.

Where a controlled entity is a very complex process involving a great number of people and equipment, and the associated control system is a complex hierarchical structure, an informational description of the system is advantageous in that it discloses the fundamental relations determining the operation of the system as an entity, the manner in which it lives.

In the informational approach, the system in question can most abstractedly be represented as a hierarchical structure (Fig. 4.8). The lowest level of this structure is occupied by the productive (process) units (shown as rectangles) which are to be controlled by the system. The higher levels are taken up by control centres (shown by circles) which are coupled to controlled units and to one another by flows of various information (shown by arrows).

With this structure, a control centre in a particular level exercises control over all the units coupled to it from below and also its own controlled units. In the general case, the same unit in a lower level can be subordinated to several units in the next higher level; that is, the structure is not a rigorous hierarchical tree.

The information circulating between the various control and controlled units within a system may take any one of three forms: input information which flows from controlled to controlling units, that is, up the hierarchical structure; control(ling) information which flows in the reverse direction; and processing information which is information about the operating algorithms prescribed for control units and their individual elements.

Primary input information originates in the various segments of the process. As it moves up the hierarchy, this information gradually takes an

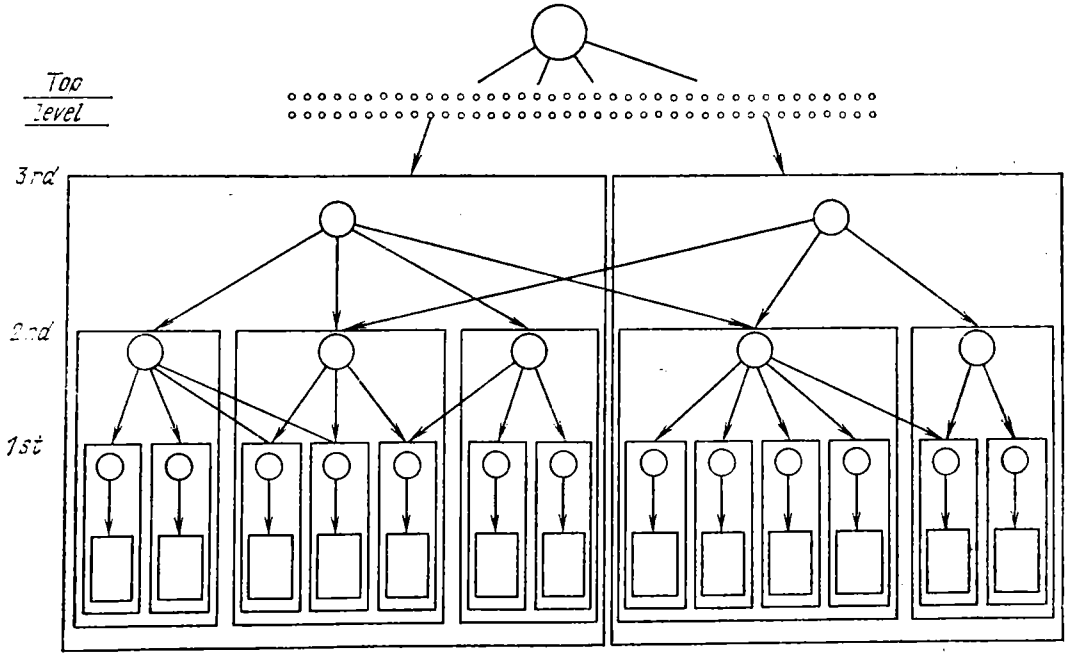


Fig. 4.8

ever more general form, is transformed in the controlling units, and arrives at the main control centre at the top of the hierarchy. Using this incoming information, the main control centre generates controlling information which flows down the hierarchical tree and is filled in with more and more detail until it reaches appropriate organizational means (final control elements).

For control goals to be attained, it is vitally important that the respective control centres receive only the necessary and sufficient information.

Table 4.1

Levels and Sections of an Informational Description

Levels of abstraction	Informational sections		
	set-up	structure	properties
System	Presentation of information from view-point of control hierarchy; goal hierarchy	Information transmission and conversion structure (communication channels and information centres). Operation programs of a system as a whole	Disorganization in system operation with regard to system goals; quantity and quality of information at inputs and outputs of information centres; generalized conversion properties and dynamic characteristics of information centres
Information centres (information generators and users), communication channels	Information presentation and conversion tasks within each centre	Same, with respect to an information centre or communication channel	Same, with respect to tasks of an information centre
Information converters (people, machines), information flows	Same, with respect to information converters. Set-up of information flows	Same, with respect to an information converter or information flow	Same, with respect to an information converter or individual programs executed by the converter

As is seen, in the operational control of a large system emphasis is placed on the qualitative properties of information, above all on its meaning and value.

An informational description of a large system should preferably be done at three levels of abstraction (Table 4.1). In analysis of very complex systems, one more level of abstraction may be added, that of subsystems.

At each of these levels, analysis can be carried out from the stand-point of that level's information set-up, information structure, and information properties.

The information set-up implies the forms in which information appears due to the hierarchical structure of control and the goal hierarchy.

The information structure, depending on the level of abstraction involved, has to do with the information transmission and processing structure on a system-wide scale, the information transmission and processing structure within each control centre, and the structure of the algorithm executed by an individual information processor.

The information properties have to do with disorganization (or organization deficiency) in the operation of the system with regard to its goals, the value of the information circulating within the system, the properties of the algorithms by which the information processors operate, the dynamic characteristics of information processing, and the dynamic properties of communication channels.

Complex processes and, the more so, large systems are always controlled under conditions of incomplete information. This is so because large systems are only partly observable, partly controllable, partly cognizable and, as a consequence, only partly predictable from the stand-point of control centres. This incompleteness of information stems from the fact that such systems are extremely complex in structure, composition and properties, on the one hand, and from the fact that it would take prohibitively much effort and time to acquire exhaustive information about (that is, to observe and get insight into) a system and to have complete control over it.

4.6. Informational Processes in Electric Power Systems

The complex information structure, hierarchical relationships between and within control centres, huge flows of information over communication channels and inside control centres, the dependence of information and its processing on goals (which fact determines the semantic and pragmatic properties of information), multiplicity of information processors widely differing in informational properties (various controllers, computers, etc., various human specialists at control centres) are features specific of all large man-made systems. On top of this, the processes occurring in electric power systems have features of their own, determined by the nature of power generation itself, by the fact that an electric power system is spread over a large area, that the generation, transmission, distribution and utilization of energy is a process continuous in time, that load varies and faults occur in a probabilistic manner, and that the quality of service, especially its reliability and security, must meet stringent requirements.

The information structure of electric power systems has taken shape over a relatively large span of time, often by adding new self-contained controllers, protection and fault-prevention means, supervisory and dispatching means to the existing control structure so as to compensate, at least in part, for the drawbacks of the past.

The situation has changed drastically since the advent of an integrated power supergrid in the Soviet Union, and a sweeping overhaul has been

started in the information structure of the system. It will culminate in a country-wide automatic management information system, already christened *Energiya* ('energy'). It is to be wholly based on systems analysis, cybernetics, wide use of computers and multiplex communication channels.

Informational structure of the integrated electric power supergrid. The control hierarchy of the supergrid may be visualized as consisting of four generalized levels each of which may in principle contain an internal hierarchy of its own. These may be the control levels of the integrated supergrid, of an interconnected electric power system, of a regional electric power system, and of a large power station (substation or network area). This hierarchy is not unlike that already adopted for the integrated electric power grid of the European USSR, but it does not preclude appropriate changes (some additions or removals) to be made in the future.

Internal hierarchy is most well-defined at the lowest control level, especially where the controlled unit is an electric power station. Here, we can differentiate routine control (circuit and connection switching, monitoring and adjustment of group and individual control systems), automatic control under normal and emergency conditions (active and reactive power dispatch, frequency control, switching control, supervision of fault-prevention means, adjustment of settings for the controllers in the lower levels); complex group governing (excitation and speed); primary self-contained governing (excitation and speed); 'zero' self-contained governing intended to maintain the energy carrier in a predetermined condition and to supply the station's own needs.

In the information structure, the control centres are, figuratively speaking, a community of operators (dispatchers), data processing and control computers, analog controllers, fault-prevention means and relays interconnected by flows of input and controlling information. At the top and middle levels, preference should of necessity be given to digital data processing means.

The rights and obligations of the members in this community are decided by the control tasks to be achieved and by the contributions made by the individual members towards the common objective. In this framework, the human operator should accomplish purposeful control on an overall strategic scale, set objectives for data processing and control machines, monitor their performance, and bear legal responsibility. The data processing and control computers should above all accumulate, store and process useful information for the purpose of forecasting, optimization of control actions, and generation of recommendations for the human operator. Group and individual control systems are to generate optimal or nearly optimal control actions to be applied over very short time intervals.

This break-down of responsibilities between the various control systems and the human operator arises from the tasks to be handled at the various levels of the control hierarchy, from the ability of man and machine to accept, store and process information according to predetermined algo-

rithms, and also from the requirements for the speed of control response and reliability of operation.

Permanent (rigid) algorithms cannot provide for an optimal operation of a complex system in a changing environment and under conditions of a changing internal structure. This can only be done by adaptive control systems. Also, control of such an entity as an electric power system will always proceed under conditions of incomplete information because the system is spread over a large area and is extremely complex in structure. The best way to make up for incompleteness of information is adaptation which may be external (through use of information coming from the higher control levels) and internal. The latter may be accomplished by adaptive systems which may in turn be of the search type which operate to a predetermined search algorithm, or of the non-search (analytical) type deriving the requisite data from a priori information and from analysis of the controlled subsystems. Under the constraints imposed on the time during which the incompleteness of information is to be made up for, important significance is attached to heuristic methods and, in particularly critical situations or following failure of automatic control systems, to the intuition (experience) of the human operator.

Structure and properties of information flows. During the operation of an electric power system, flows of input and control information are circulating between the control centres.

Input information includes messages on the status and parameters of the system elements and of the system as a whole (the present coordinates of the system state), and various technical and economic indicators of system performance. Control information includes orders or instructions to generate so much energy, to change the status of particular system elements or state variables, and also instructions to change the algorithm of a particular control centre.

Input and control information may be generated and utilized both inside and outside the control system and make up flows of information connecting the control system with the surroundings.

Figure 4.9 shows the jth control centre in the ith level and all the incoming and outgoing flows of information. The control centre accepts input information $I_{in\Sigma}^{i-1}$ from all adjoining control centres in the $(i-1)$st control level, external input information I_p^{ext}, external goal-oriented control information $I_{\alpha(g)}^{\text{ext}}$, and goal-oriented control information $I_{\alpha(g)}^{i+1}$ from the control centre in the $(i+1)$st control level. After all of this information is processed under the action of processing information I_L embodied in the algorithm, the output of the control centre delivers flows of control information $I_{c\Sigma}^{i}$ for the centres of the $(i-1)$st level, input information I_{in}^{i} for the centres of the $(i+1)$st level, and input information $I_{in.\,ext}^{i}$ for external systems. Control centres in the $(i+1)$st level can also supply control information I_{cL}^{i+1} to change the algorithm of the jth control centre, and input information I_{in}^{i+1} which cannot be derived in the ith level but is essential to decision-making at the jth centre. Thus, in the general case, control informa-

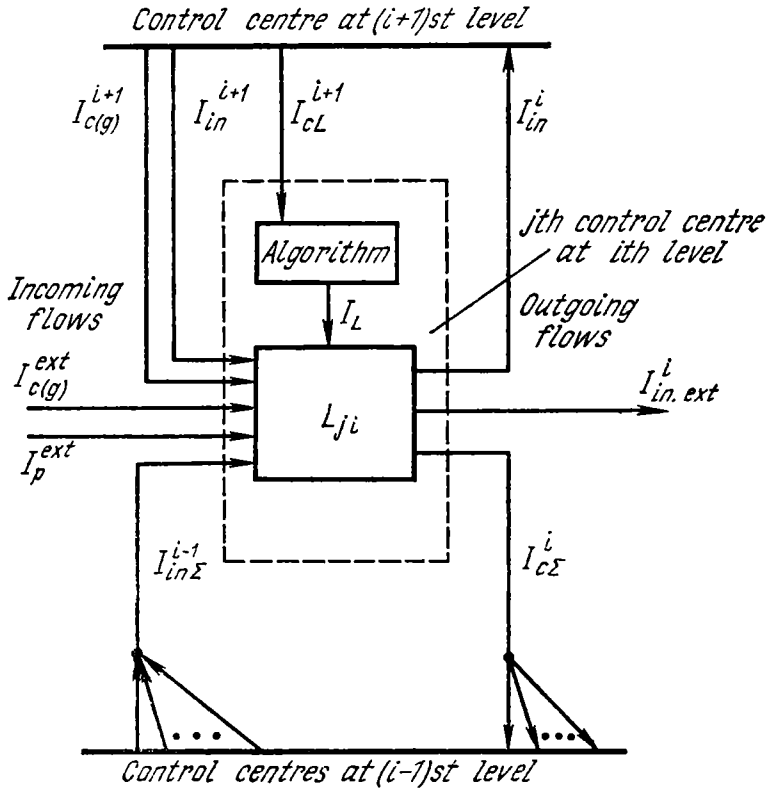


Fig. 4.9

tion may be of two types, goal-oriented and algorithm-defining or modifying.

Because, practically, the surroundings for the jth centre may be control centres in the same ith level, input information can be transferred from the higher to the lower centres, and also there may be cross-couplings between the centres, we may conclude that the informational structure of an electric power system is, strictly speaking, a hierarchical graph which reduces to a hierarchical tree in a special case only.

By the manner in which it is utilized, all information making up internal and external flows may be divided into two large classes:

(1) information utilized for the operational control of the power system (operational information);

(2) information utilized to optimize the technico-economic performance of the power system (technico-economic information).

According to the system state in which it is generated and utilized, operational information may be classed into three groups as follows:

(1) information for the operational control of the system in the normal state;

(2) information for the operational control of the system in the abnormal (or emergency) state;

(3) information needed to optimize the system performance in the course of system control.

In the normal state, operational information is utilized to maintain quantitative and qualitative performance indicators (power generation and transmission, load voltage, line frequency).

In the abnormal (emergency) state, instead of maintaining the specified performance level one seeks to preserve the system stability, and to minimize the cost of loss due to energy not supplied because of interruptions in service.

In the case of optimization, operational information is utilized to determine optimal quantitative indicators which are to be maintained in the normal state of the system. In such a case, an economic criterion is optimized, subject to constraints on the security of service and the quality of power supply.

The following flows of information are included in each of the above groups at any control level:

(a) data related to the electric parameters of system elements or to the electrical part of system operation;

(b) data related to the thermodynamic and hydrodynamic parameters of system elements or system operation;

(c) data related to parameters which affect system performance indirectly;

(d) data related to the status of elements in the power and control systems.

Consider each group separately.

Flow 'a' will typically be concerned with system frequency, voltages at test points, currents in the main branches, the active and reactive power supplied by generating units, stations and systems, active, reactive and assist power transmitted over the lines, the reactive power supplied by synchronous condensers, the resistances and conductances of the power transmission line, transformation ratios, the reactances of generators, lines, and synchronous condensers, the time constants of excitation control systems, etc.

Flow 'b' will typically include thermal power parameters, such as steam, oil and water pressures, fuel and water rate of consumption, water tank levels, conductivity of water in the boiler and condenser, the concentration of hydrogen ions in steam-generating water, the temperature of the oil cooler, steam superheater and preheater, the temperature of bearings, etc.

In the case of nuclear power stations, additional data will be included on the distribution of neutrons across the radius of the active core of the reactor, consumption rate of nuclear fuel and flow rate of coolant, the temperature at test points of the reactor and of coolant, etc.

For a hydro station, it is important to know the flow rate of water, water storage, head water level, etc.

For any type of electric station, it is important to know the rotational speed of turbines, torques on turbine shafts, the parameters of boilers, water and steam turbines, etc.

Flow 'c' covers weather conditions in the various areas of the power system, weather and hydrologic forecasts, astronomical and synchronous time, etc.

Flow 'd' includes data on the configuration of transmission and distribution networks, the status of system elements (lines, transformers, rectifiers, etc.), generating units and stations, the settings and adjustments of controllers and relays, the settings of the under-load tap changers of transformers, the positions of valves, etc.

Information in different flows is generated at different rates. The approximate upper bounds on the value ranges of parameters in various system states are listed in Table 4.2.

Since the value range for any system or state variable is bounded, while the number of controlled variables is great, operational control systems mainly change the variables at discrete time intervals (except some automatic control and protection systems), and such systems are called sampled-data systems.

According to the instants when variables are measured or sampled, sampled-data systems may be classed into programmed and sporadic. In the former case, information is generated either at predetermined instants or

Table 4.2

Bounds on Value Ranges of Parameters

Information flow	System state			Value range of parameter
	abnormal	normal	optimization	
Flow 'a': electrical parameters	10^2	10	10^{-1}	Upper bound on value range, Hz
Flow 'b': thermal and hydrodynamic parameters	10	1	10^{-3}	Same
Flow 'd': element status	10	10	10	Maximum time, μs, for elements to change state, without account of using this information in control system

in a predetermined sequence by several sources. In the latter case, information is generated at some random instants selected either by a source or by the user.

The choice of a particular system and the associated matters of accuracy and delays are mainly dictated by the purpose served by the users, that is control systems, and by the informative content of data from the standpoint of pragmatic utility and attainment of control goals.

Sample informational analysis of the structure and set-up of an operational control system for a hydraulic power station under normal conditions*. Power generation by an electric station may be divided into a number of stages each involving a specific mix of generating plant and processes, as follows:

(a) hydraulic energy is converted into electricity by the generating units of the station, including the generating units supplying the station's house needs;

(b) electric energy is distributed among the station's consumers through the primary connections, including main transformers;

(c) electric energy is distributed among the station's own loads through the station's house connections, including house transformers;

(d) electric energy is converted for utilization within the station.

The above division into stages, units and processes is rather arbitrary, but it is useful in gaining insight into the control of a hydro station. On the basis of such division the operational control system of the station may be depicted in block-diagram form as shown in Fig. 4.10.

The lowermost level of the block diagram (which may arbitrarily be called the zero level) is occupied by the controlled units which include eight hydraulic generating sets GS_1 through GS_8, the station's house sets, the main power connections, and the station's house connections. Above each of these units is its control system (or control block) CS_1 through CS_8, made up of a variety of controllers, regulators and relays. The control units are coupled to the controlled units by flows of input information directed from bottom upwards, and of control information flowing from top to bottom. These units form the first level of the control hierarchy.

The next, second, level is occupied by a group control system or block, for which the controlled units are the generating sets together with their own control units. Flows of information couple this control system to the generating sets proper and to their controls.

The top, third, level of hierarchy is occupied by operators, *OP*. The team of operators utilize all operational information about the station's performance and initiate the necessary corrective or control actions, when and if necessary. Within this team, duties and subordination are prescribed in a precise and mandatory manner.

The block diagram of Fig. 4.10 is only an outline representation of a control system, solely intended to illustrate the hierarchy of operational control within a power station. A detailed diagram would show about 100 control nodes.

In the general case, each control centre utilizes a set of incoming flows of information and generates a set of outgoing flows. The flows of information entering a control centre may be of any one of three types:

(1) input information objectively describing the properties, set-up and structure of the controlled units and the environments;

* After B. A. Orekhov who has conducted an informational study at a hydro station.

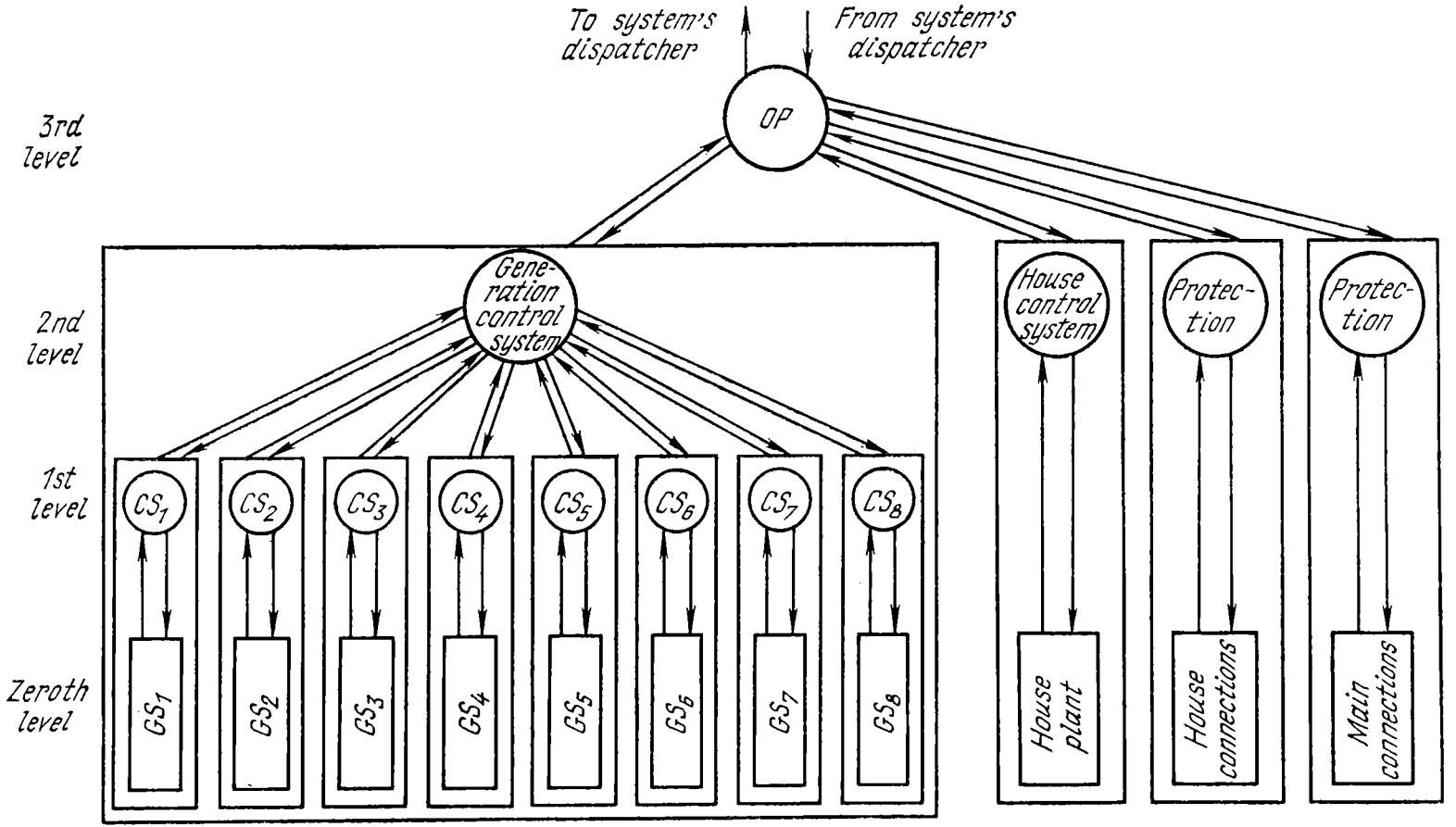

To system's dispatcher
From system's dispatcher
OP
3rd level
2nd level
1st level
Zeroth level
Generation control system
House control system
Protection
Protection
CS1
CS2
CS3
CS4
CS5
CS6
CS7
CS8
GS1
GS2
GS3
GS4
GS5
GS6
GS7
GS8
House plant
House connections
Main connections

Fig. 4.10

(2) control information coming from the higher levels of control. It consists of orders and instructions with regard to the process variables under control of the centre. In other words, this flow carries information about the control goals of the respective centres and the values of the controlled variables that are to be maintained for normal operation of the entire system;

(3) processing information which is in effect likewise control information and selects algorithms for the processing of the first two incoming flows into outgoing ones. Processing information may arrive at a control centre either from higher levels of the operational hierarchy or from centres which are not part of the operational hierarchy directly, but belong to a more general control hierarchy. Besides, processing information may be embodied in the structure of a control centre, as is the case with simple automatic controllers operating to a rigid algorithm, or it may come as information about an adaptation algorithm for automatic controllers; or it may be accumulated by a control centre in the course of learning, if the centre has capabilities for filtering and storing useful information.

The configuration of the communication channels interconnecting the control centres, together with the information transferred, forms the informational structure of an electric station.

Each flow of information is conveyed over a communication channel. In the system under consideration these channels may be various devices for the transmission of continuous or pulse signals, telephone or telemetry lines, the measuring circuits of instruments, or secondary switching circuits. Also, information may reach the control centres in the form of visual, aural and other patterns or images.

A flow of information has several constituents. A feature common to all constituents is that they have the same source and the same sink.

A description of the information carried by a flow may be given as a table which lists:

(a) units of measure;

(b) the likely range of values, that is, the values that a variable may take in accordance with the physical laws governing the system;

(c) the safe range of values, that is, the values that a variable may take without endangering the safety of the system (besides, there is an emergency range which encompasses the values that a variable may take under emergency conditions);

(d) an optimal range, that is, the values falling within part of the safe range and chosen as goal, or to be maintained at all times;

(e) the least discernible value, Δ, or the quantizing (or sampling) step, below which it has no sense to measure changes in a variable;

(f) the total number of values that a variable may take or the total number of quantizing (or sampling) steps within the likely range of values:

$$N = D/\Delta$$

where D is the likely span of values;

(g) the maximum amount of information yielded by a single observation or measurement of a variable. According to the Hartley law, it is expressed as

$$I = m \log_2 N$$

where m is the number of identical observables.

For each constituent of an information flow the table defines the manner in which it is to be observed or measured and the rate at which it is to arrive at a control centre. It is to be noted that the rate of arrival has sense only when control is accomplished under conditions free from abnormal situations or routine changes in system performance. For in such circumstances the control system acquires information under a different program, and both the rate and manner of observation are changed. Four types of observation or measurement can be named: continuous monitoring, periodic scanning, sporadic scanning at instants when the variable in question moves beyond the limits of the operating range, and sporadic scanning at instants selected by the control centre (on-demand scanning).

Continuous monitoring is ordinarily accomplished by simple automatic controllers and protective devices. When the variable under observation moves beyond a preset limit, they send a warning signal to a higher level (or levels) and generate a corrective action to be applied to the process. These devices are highly specialized — they are capable of accepting information limited in meaning. As is seen, they operate at variable time instants determined by their settings.

Periodic and sporadic scanning is usually accomplished by elements of a higher type (man or a computer). The action is periodic or sporadic because they utilize a considerable number of information flows which they cannot handle all at the same time. Corrective actions are likewise generated at discrete instants of time.

In the example under consideration the investigator has identified 122 flows of information. As an illustration, Table 4.3 gives fragments from two information flows.

Information estimates have shown that under normal operating conditions the control centres receive information at the total rate of 7080 bits/s. This figure has been computed on the assumption that automatic devices operating continuously take in data at a frequency $f = 100\ s^{-1}$, because information about alternating current to which these devices respond is discrete. The information load on operating personnel is 0.3 bit/s. The system's dispatcher receives 0.01 bit/s. The total amount of control information circulating in the system under consideration is 1300 bits, and the amount of information generated by the main control centre is 900 bits. Since practically all control information is generated sporadically, the above estimates are only approximate.

In accordance with relevant codes, the overall task facing the control system of an electric station may be stated as follows: maintain the conti-

Table 4.3

Information Variables

Data carried in each flow	Unit of measure	Range of values			Least discernible value	Number of quantizing levels	Maximum amount of information in one observation	Method of control and frequency
		likely	safe	optimal				
Flow I								
Position of doors (two doors)	m	0-5.5	4.5-5.5	5.3-5.5	0.05	110	13.5	Every 4 hours
Pressure in hydraulic gate jacks (two jacks)	atm	0-110	60-110	80-110	5	22	8.9	Every 4 hours
Upstream water level	m above sea level	452-457	454-457	455-456	0.01	500	8.9	Every 2 hours
Downstream water level	m above sea level	425-430	426-429	426-429	0.01	500	8.9	Every 2 hours
Flow II								
Rotational speed of generating unit (for synchronization)	% of n_{nom}	0-180	30-140	90-105	5%	36	6.45	Sporadic
Generating unit power:								
active, P	MW	0-82.8	0-82.8	$P \pm 05$	0.1	828	9.65	Every hour
reactive, Q	MVAr	0-62.8	0-62.8	0-62.8	0.1	628	9.25	Every hour

nuity of electricity supply of the requisite quality and in the requisite amount and run the station at the required level of economy.

As is seen, this task has four goals to achieve:

(a) to maintain a specified level of security (continuity) of supply or service;

(b) to generate electric energy of a specified quality;

(c) to generate a specified quantity of electric energy (to meet the generation plan);

(d) to run the station at a specified level of economy.

The four indicators appearing as the goals of control are integral quantities describing operation of the system over a fairly long time period (month, year, five-year period) falling under the category of economy or long-term management. These goals may be regarded as global for the control system of an electric station.

The operational control task is more narrow and consists in maintaining the output parameters of the station and the performance of the plant within the limits fixed by the plan, that is, in making up for various departures in the performance of the plant from that envisaged in the plan.

Thus, operational control is, in fact, an interpretation of the overall task facing the whole of the control system and formulated for a relatively long time period; it is interpreted so that it reduces to the control of the current system state.

From analysis of operation of the control centres at an electric station under normal conditions, it is possible to construct a goal tree which has four levels, seven goals and 121 subgoals. The tree is constructed on the assumption that the station is operating under conditions of total active power regulation. In this tree, there is a specific control task associated with each goal, and each goal is described in a formalized manner. A fragment of this tree appears in Fig. 4.11. Our discussion will be limited to goals *a* and *e* and subgoals *aa* and *al* in this fragment.

Goal a. The objective variable is the total active power available, $P_{\Sigma avail}$, which may take values in the range from 0 to 660 MW. The control task includes the requirement to run the station so that its total power be as close as practicable to the maximum likely value determined with allowance for scheduled plant outages. For example, the goal may be taken as attained, if $P_{\Sigma avail}=660$ MW, in which case all the eight generating sets installed at the station are operating. This value falls within the well-ordered domain*. The goal is only partly attained if

$$P_{\Sigma avail} \geqslant P_{\Sigma avail.s}$$

where $P_{\Sigma avail.s}$ is the specified value of available power, which is a multiple of the nominal capacity of one generating set, P_{nom}, or $P_{\Sigma avail.s}=nP_{nom}$ (where n is the number of generating sets running).

* For definition, see Sec. 4.8.

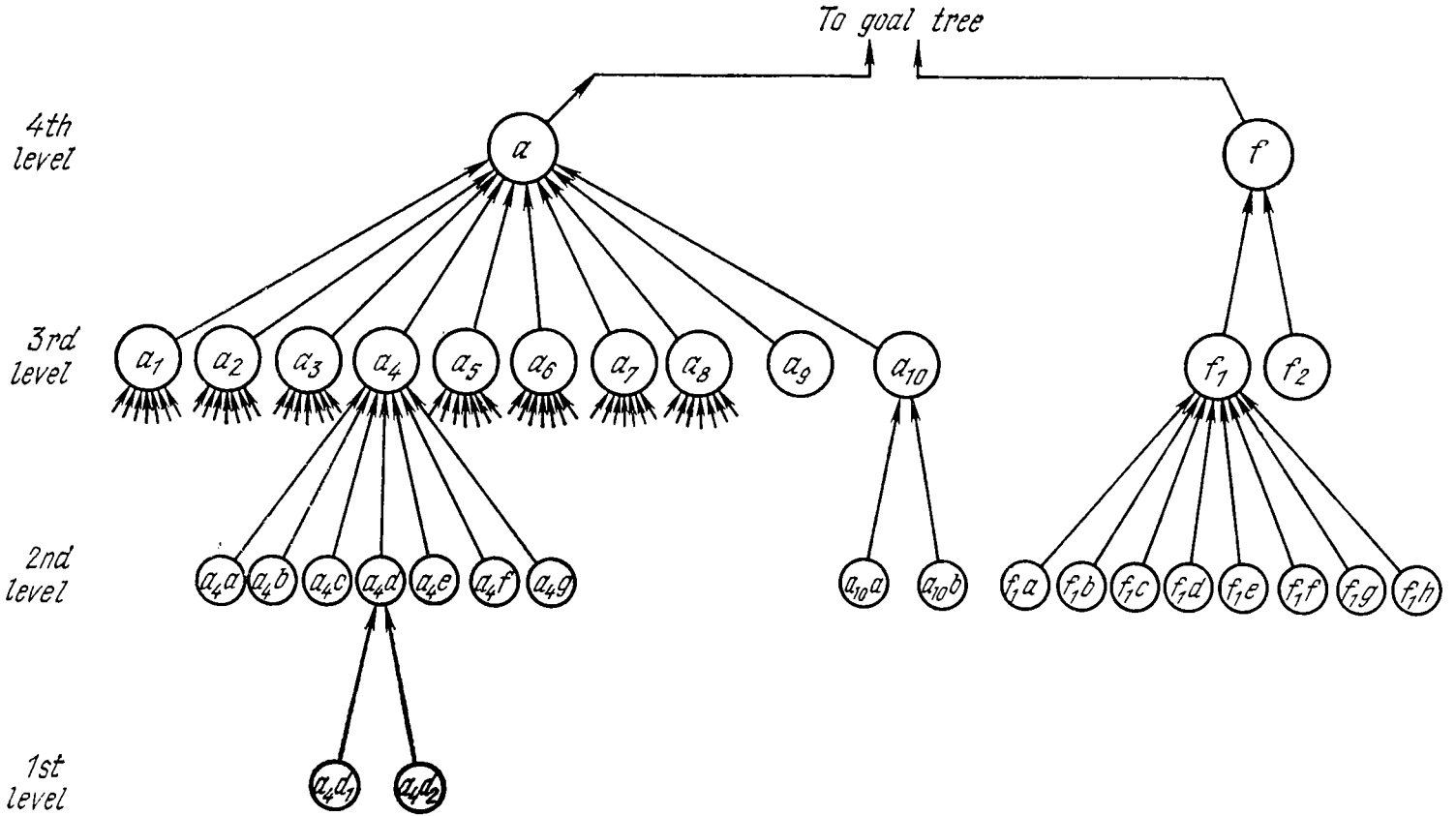

To goal tree
4th level
3rd level
2nd level
1st level
a
f
a1
a2
a3
a4
a5
a6
a7
a8
a9
a10
f1
f2
a4a
a4b
a4c
a4d
a4e
a4f
a4g
a10a
a10b
f1a
f1b
f1c
f1d
f1e
f1f
f1g
f1h
a4a1
a4a2

Fig. 4.11

All values of $P_{\Sigma avail.\,s}$ make up a partial-order domain. The goal is not attained altogether if

$$P_{\Sigma avail} < P_{\Sigma avail.\,s}$$

Goal e. The goal space has two coordinates, one being the voltage across the 110-kV bus system, V_{110}, and the other, the voltage across the 220-kV bus system, V_{220}, and the system in regard to this goal is described by a vector, $\dot{V}=\{V_{110}, V_{220}\}$. The control task is to maintain the specified level of voltage across both bus systems.

Graphically, the well-ordered domain may be visualized as a circle with the point $(V_{110(s)}, V_{220(s)})$ as centre, which is the specified magnitude of the vector $\dot{V}$, and of radius $\xi=1$ kV.

The degree of disorder is

$$\overline{Y}_V=\frac{\sqrt{V_{110}^2+\Delta V_{220}^2}-\xi}{\xi}$$

where $\Delta V_{110}=V_{110}-V_{110(s)}$ and $\Delta V_{220}=V_{220}-V_{220(s)}$.

The safety limit* which bounds the goal domain is a circle with a radius of 5 kV.

Subgoal $a_i a$ $(i=1, 2, \ldots, 8)$. Each generating set has two doors, one fast-acting and the other slow-acting. Therefore, the space of subgoal variables is two-dimensional: $\dot{a}=\{a_f, a_s\}$, where a is the position of the door, in metres, measured from the closed (lowermost) position. The control task in this case is to hold the door in the specified position. The well-ordered domain for each coordinate lies in the range 5.32 m $< a \leqslant$ 5.5 m. The safety limit occurs at $a=4.5$ m. The point $a=5.32$ m represents the door position in which an alarm is actuated.

Subgoal $a_i l$. The task is to maintain a specified water pressure difference across the turbine bearing. The subgoal parameter is the pressure differential $\Delta D_{t.b.}$ across the turbine bearing. The well-ordered domain extends from 0.3 to 1.0 atm. The safety limit is the same as the boundary of the well-ordered domain.

4.7. Trends in the Development of Information Theory

It should be admitted that the application of informational analysis to power-system problems has so far been limited by a qualitative description of the control structure, the composition and structure of information flows and the quantitative distribution of their constituents, such as the rate of data transmission over communication channels and the amount of information associated with a particular message.

The existing body of information theory is not capable of doing the main thing — to determine the pragmatic and semantic utility of information

* To be defined in Sec. 4.8.

with regard to control goals. The foregoing applies to any other system as well as to problems of power system control.

The cause lies in the fact that the techniques employed by the statistical theory of information have proved unsuitable for the analysis of control problems, although the latter have given big impetus to the development of cybernetics as a science. The statistical theory of information is very powerful in describing the processes involved in the transmission and storage of information, but it is ineffectual when it comes to treating information in relation to control goals.

At present, it is urgent from the view-point of cybernetics as a whole and of its divisions, notably the theory of large systems, especially large power systems, to develop a new body of methods to describe the processing of information with special reference to control tasks. In other words, what is needed is a new information theory, the *information theory of control.* It appears that such a theory should meet the following requirements:

(1) with regard to the principles used, it must be analogous to the classical theory of automatic control, but it must be raised to a higher level of abstraction in the sense that the concern should be with the information carried by signals rather than with the signals themselves. In this approach, instead of the usual transfer functions, automatic control means should be characterized in terms of informational functions reflecting the properties of information processing algorithms with regard to the attainment of a particular goal, and also in terms of informational efficiency (or performance) functions which assess how a given information processing algorithm is realized by a particular control device (subject to constraints on, say, speed of response, storage capacity, range of numbers, etc.);

(2) the principles by which the informational characteristics of systems can be derived experimentally should be similar to those of the frequency domain analysis so widely employed in the theory of automatic control.

As a result, a higher level of abstraction can be achieved in systems description and analysis. This is what is needed in the investigation of large systems as integral entities where classical methods fail because the results obtained are extremely complex and difficult to generalize.

The informational theory of control is still in embryo, and before we take up its main points, it will be worth while dwelling on the general trends in its development.

The existing forms and measures of information have been classified in the literature with sufficient completeness. Most often, use is made of the statistical measure of information. As already noted, however, the statistical theory of information has proved ineffectual in the analysis of control processes and the description of system operation. This is because statistical information describes only the quantitative aspect of a message but ignores its qualitative aspects, that is, its utility, contribution to the attainment of an objective, and its relation to the receiver's level of enlightment or skills. Furthermore, any system operates and expands in a purposeful manner determined in the general case by a dynamic goal tree. The inade-

quacy of the statistical measure of information in characterizing the various aspects of informational processes in systems has been conducive to the adoption of nonstatistical measures of information (topological, combinatorial, algorithmic, dynamic), and also to quantitative interpretations of information.

Apart from the continuing effort to differentiate various aspects of information, attempts have of late become increasingly often to generalize the ideas and methods of information on the basis of the most basic concepts — those of variety and degree of organization. It may be expected that further advances in the field of informational methods will finally lead to a general information theory. Besides, to meet engineering, biological and sociological needs (including those of power engineering), a number of applied information theories may well appear, based on the informational theory of control and the informational theory of modelling and programming.

As already noted, the most general concept in information theory today is variety. Although this concept was advanced by W. R. Ashby (in his Law of Requisite Variety) as an equivalent of the statistical theory of information, it has become an embracing one in a philosophical plane for all existing forms of information presentation, both quantitative and qualitative, only recently. Today, the theory of information is attempting to replace the simpler representations of information, associated with its quantitative forms, by the more complex qualitative forms. Among the factors that are promoting this tendency are physiological, psychological and biological concepts of information and further advances in the qualitative aspects of its theory. In this connection, special interest is evoked by the assessment of the utility (value) of information in terms of indicators describing the rapidity with which a goal is attained.

A very important direction in information theory is work on the semiotic theory of information which embraces syntactics, relating the formal properties of characters and their combinations to the amount of information they carry, and semantics and pragmatics, relating the characters to information content and information utility, respectively.

At present, a new (organizational) direction has appeared in information theory, which relates information to orderliness and degree of organization. This trend can be traced back to N. Wiener and W. R. Ashby, although more or less precise definitions and concepts, even in a philosophical plane, have been formulated only recently. In its development, this trend will inevitably cooperate with semiotics to promote the advent of a general information theory.

Engineering cybernetics has often to deal with problems which cannot be tackled by the classical methods of control theory. This is why it is particularly attractive to approach them from the stand-point of information theory, as it enables such problems to be taken up at higher levels of abstraction. A very important element of this informational theory of control is based on *the threshold of distinguishability and the amount of variety*. Like

the information theory of communication, the information theory of control is a pure theory independent of the form of the object under examination. Its capabilities are so far limited to defining the limiting characteristics of automatic control systems and an informational description of some aspects of the control process. Since it is still detached from control objectives, this theory remains purely quantitative, isolated from pragmatics and semantics which represent the qualitative aspect of the control process. We have already listed the requirements that an informational theory of control must satisfy in order to be useful in tackling the basic problems of control. Some possibilities for developing an informational theory of control will be taken up in Sec. 4.8.

Analysis and synthesis of man-machine systems, analysis of biological systems, and synthesis of optimal information machine systems make it urgent to develop a special information theory — an information theory of modelling and programming. Work done to date in this field has been limited to behavioural models (or the information structure level, in the terminology of Table 4.1) and a descriptive characterization of situations. No consideration has been given to the information content of behavioural programs or to the pragmatic value (utility) of such programs to the user of information, be it man or a machine. An information theory of modelling and programming is especially needed for an informational description of large systems. It is imperative to set up a number of similarity criteria with regard to goals, information structures, information flows (in terms of the amount and value of information), and to establish similarity between informational functions of information processing by control centres. This latter problem arises from the need to assess the operation of the programs executed by digital computers which act as sources of useful information for the human operator incorporated in a man-machine system. It may be expected that big strides will be made in the information theory of modelling and programming before long. Today, however, this ground has only been probed by a few investigators.

Figure 4.12 shows a tree which depicts the trends in the development of informational methods and their relation to other theories (existing theories and conceptions are shown on the right of the diagram, while new theories in the making and related tasks appear on the left).

When it comes about, the applied information theory of large power systems will undoubtedly give big impetus to the progress of the power industry. It is not unlikely that some of the readers will contribute to such a theory in their practical work. The road to the information theory of power engineering lies through the information theory of control and the information theory of modelling and programming which will be common to all systems.

However, such theories are non-existent yet, although work is under way and some of the problems are already being tackled. We shall dwell on them in brief, placing main emphasis on the methodology and the body of mathematics used.

4.8. Brief Introduction to the Information Theory of Control

We shall examine the points of departure for such a theory in accordance with Fig. 4.12, and establish relationships between complexity, orderliness, degree of organization and entropy.

With regard to system mix, complexity may be a function of any one of two basic variables, the number of elements that make up the system or their variety. Similarly, with regard to structure, system complexity may be a function of the number or variety of flows or connections, and with regard to system properties it may be a function of the number or variety of properties. For example, the algorithmic complexity of a system (the complexity of its behaviour) may be expressed as a function of the number or variety of elementary operations included in an algorithm. At a higher level, it may be appropriate to speak of algorithmic complexity as a function of the total number or variety of algorithms available to a system.

Also, the above complexity variables may be examined in space or in time. For example, algorithmic complexity in time may be expressed as a function of the number or variety of the algorithms utilized by a system to shape its behaviour over a selected time interval, while algorithmic complexity in space as a function of the number or variety of the algorithms stored in the system. In the light of the foregoing, complexity may be defined as follows.

Complexity is a generalized estimate, from an observer's stand-point, of the number or variety of system elements: mix (composition), structure and properties, considered in space or in time, or both. In the general case, the system conponents are related to complexity non-additively and are a function φ.

The complexity C of a system may then be defined as

$$C_{in}=\varphi(n) \tag{4.19}$$

$$C_{iN}=\varphi(N) \tag{4.20}$$

where n is the number of elements considered and N is their variety.

The functional relation φ may in most cases be approximated by four basic relations, linear, power, exponential and logarithmic. In the last case, expressions of the form $\log n$ or $\log N$ will be a kind of indicators of system complexity, and numerically they may in some cases be equal to the amount of information defined by Hartley's law.

Orderliness is a characteristic of order, that is, conformity with something adopted as a standard. A standard of order may be a specific arrangement or ranking of elements within the system under consideration or a specific relation between two or more entities. In all of these cases, it is more convenient to employ the reciprocal of orderliness, that is, disorderliness.

Disorderliness is a measure of difference of a selected variable, say x, with regard to a standard of order, x_s, this difference tending to zero when $x_j \to x_s$. Here, x_j is the variable under consideration (it may be a position,

time, statistic or any other variable) in the jth realization, and x_s is the standard value of the same variable assumed to be the centre of orderliness. Around this centre, there is a quasi-ordered region inside which order is maintained accurate to ε. Accordingly, disorderliness ought to be assessed in terms of $\xi(x)$, the distance from the centre of orderliness to the boundary of the quasi-ordered region.

According to the form of reference used, disorderliness may be:

(1) absolute

$$\bar{Y}_{ja}=|x_j-x_s|-\xi(x) \tag{4.21}$$

Since we are dealing with a vector quantity, Eq. (4.21) may be re-written in the more general form as

$$\bar{Y}_{ja}=q_j-\xi(x) \tag{4.22}$$

where q_j is the magnitude of the error vector;

(2) fractional

$$\bar{Y}_{jf}=\bar{Y}_{ja}/\Delta \tag{4.23}$$

where Δ is a quantity in terms of which absolute disorderliness is weighted. For example, Δ may be a quantizing level or any region;

(3) power

$$\left.\begin{array}{lll} \text{(a)} & \bar{Y}_{jp}=\dfrac{x_s-\xi(x)}{x_j}-C & \text{for} \quad x_s>x_j \\ \text{(b)} & \bar{Y}_{jp}=\dfrac{x_j}{x_s+\xi(x)}-C & \text{for} \quad x_s<x_j \\ \text{(c)} & \bar{Y}_{jp}=\dfrac{q_j}{\xi(x)}-C & \end{array}\right\} \tag{4.24}$$

We assume the system under consideration to be well-ordered, that is, $\bar{Y}_j=0$ if $q_j-\xi(x)\leqslant 0$ or $|x_j-x_s|-\xi(x)\leqslant 0$. The attendant requirements are that in Eqs. (4.22) through (4.24) the choice should always be such that $q_j\geqslant\xi(x)$ or $|x_j-x_s|\geqslant\xi(x)$, while in Eqs. (4.24) C should always be chosen such that $\bar{Y}_j=0$ for $q_j=\xi(x)$ or $|x_j-x_s|=\xi(x)$. At the same time, C may be constant ($C=1$) or varying, for example, $C=\exp(\bar{Y}_{ja})$.

Disorderliness may also be assessed with regard to objective or subjective goals, some region of interest (for example, an uncertainty region), an external reference system, the ranking adopted as a standard, etc. By expressing disorderliness as

$$\bar{Y}_j=\varphi[x_j, x_s, \xi(x)]$$

that is, as the difference (diversion) referred to something, it may be:

(1) deterministic, in which case x will likewise be deterministic;

(2) probabilistic, as a result of x being a probabilistic quantity, say, variance;

(3) an uncertainty indicator, where two cases may occur:

(a) x gives a measure of the uncertainty region (deterministic or probabilistic);

(b) x defines the a priori probability of a sample (the uncertainty region or the uncertainty of a sample within the region).

Like complexity and orderliness, degree of organization may be examined with regard to system mix, structure and properties.

When considered with regard to control goals, degree of organization is a function of both the amount of useful information contained in the control program, and the capacity of organizational means which implement this program as actions directed to attain the control goals.

Like orderliness, degree of organization can more conveniently be measured in terms of its reciprocal, organization deficiency or disorganization.

Organization deficiency or *disorganization* is a generalized disorder characteristic of some system mix, structure or property indicators, weighted in terms of its relevance to the variables of the higher system levels (say, to the attainment of control goals). It can be evaluated over a number of elements, situations or time intervals.

Thus, disorganization is a derivative of disorderliness and complexity. Here, complexity manifests itself indirectly as the number of elements, situations or time intervals.

By definition, the following expression may be written for disorganization:

$$\bar{O}=\bigcup_{\beta=1}^{l}\alpha_\beta\bigcup_{i=1}^{d}s_i\bigcup_{j=1}^{m}p_jf(\overline{Y}_j) \tag{4.25}$$

where $f(\overline{Y}_j)$ = disorderliness weighted in terms of its relevance to the jth situation

$\bigcup$ = symbol of union over l time intervals, d elements and m situations, respectively

α_β, p_j, s_i = weights of the βth time interval, dth element and jth situation, respectively

In principle, we may speak of unit disorganization, that is, the disorganization per situation, per element or per time interval:

$$\bar{O}_{unit}=f(\overline{Y}) \tag{4.26}$$

If there are a total of m situations, and the probability of occurrence of each jth situation is p_j, then within the ensemble of situations

$$\begin{pmatrix}\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_m\\ p_1, p_2, \ldots, p_m\\ \omega_1, \omega_2, \ldots, \omega_m\end{pmatrix}$$

the generalized organization deficiency, or disorganization, may be defined as

$$\bar{O}=\sum_{j=1}^{m}p_j\psi(\Pi_{\bar{y}ja}) \tag{4.27}$$

where $\Pi_{\bar{y}ja}$ is the disorderliness parameter defined as

$$\Pi_{\bar{y}ja}=\bar{Y}_{ja}\omega_a+C \tag{4.28}$$

or as

$$\begin{aligned}\Pi_{\bar{y}jf}&=\bar{Y}_{jf}\omega_f+C\\ \Pi_{\bar{y}jp}&=(1/\omega_p)(\bar{Y}_{jp}+C)\end{aligned} \tag{4.29}$$

where ω_j = weight taking care of the qualitative difference of Y in the jth situation from other situations (with the entire ensemble having a common relation ψ)

C = coefficient defined from the condition: for $\bar{Y}_j=0$, $\psi(\Pi_{yj})=0$

Thus, as follows from Eqs. (4.27) through (4.29), disorganization is an ensemble-generalized characteristic of disorder in the state or behaviour of a system, weighted firstly in terms of situation probability and secondly in terms of its relevance. The relevance of disorder to the jth situation is decided both by the relation ψ common to the entire ensemble and the additional weight ω_j taking care of the difference in $\bar{Y}_j$ between the jth and other situations.

The disorder parameter may have a more complex form if it is dependent on several constituents. For example, the disorder parameter for the jth situation in the case of m components represented in a metric space by the expected values of absolute disorder and the weights ω_{ij} will be defined as

$$\Pi_{\bar{y}j}=\sqrt{\sum_{i=1}^{m}(M\{\bar{Y}_{ja}\}\omega_{ij})^2}+C \tag{4.30}$$

For most systems, the relation ψ will be linear, power, logarithmic or exponential.

As will be shown later, in control problems the relation ψ may be interpreted as the sensitivity of a higher-level goal towards the degree of failure to attain the subgoals considered. On the basis of Eq. (4.27) and the assumptions made with regard to ψ, the following basic expressions may be written for disorganization:

$$\bar{O}_l=k\sum_{j=1}^{m}p_j\Pi_{\bar{y}j} \tag{4.31}$$

$$\bar{O}_p=\sum_{j=1}^{m}p_j(\Pi_{\bar{y}j})^n \tag{4.32}$$

$$\bar{O}_{log}=\sum_{j=1}^{m}p_j\log\Pi_{\bar{y}j} \tag{4.33}$$

$$\bar{O}_{exp}=\sum_{j=1}^{m}p_j[\exp(\Pi_{\bar{y}j})-C] \tag{4.34}$$

where C, k and n are the constant coefficients and the power relation specified in advance.

Assuming that disorganization with regard to elements and time is additive, that is, assuming that the components of these disorganizations can be added together on the basis of Eq. (4.25), we may write the following general expression for disorganization:

$$\bar{O}_{\Sigma}=\sum_{\beta=1}^{l}\alpha_{\beta}\sum_{i=1}^{d}s_{i}\sum_{j=1}^{m}p_{j}\psi(\Pi_{\overline{yj}}) \tag{4.35}$$

To be able to determine the organization deficiency of a system from those of its parts and to carry out operations on disorganizations, it is necessary to set up rules for their addition and subtraction, which would be valid with sufficient accuracy for both additive and non-additive forms of disorganization.

If, for a specified relation ψ, we have evaluated the generalized disorganization, $\bar{O}$, then using the inverse function, ψ^{-1}, we can evaluate the generalized disorder parameter (for the whole ensemble) corresponding to this disorganization as

$$\Pi_{\bar{y}}=\psi^{-1}(\bar{O}) \tag{4.36}$$

Hence, the rules for the addition (union) and subtraction (difference) of disorganizations accurate up to constants C may be formulated as follows*.

To obtain the union of n independent disorganizations, it is necessary first to obtain their images on the disorder parameter set, add (or multiply) together the values of these images, and map the result back onto the disorganization set, that is,

$$\bigcup_{j=1}\bar{O}_{j}=\psi\left\{\sum_{j=1}^{n}[\psi^{-1}(\bar{O}_{j})-C]+C\right\} \tag{4.37}$$

or

$$\bigcup_{j=1}\bar{O}_{j}=\psi\left\{\prod_{j=1}^{n}[\psi^{-1}(\bar{O}_{j})-C_{j}]+C\right\} \tag{4.38}$$

where $C_{j}=C/\omega_{jp}$.

Similarly, to obtain the quantitative difference between (or intersection of) disorganizations, say $\bar{O}_{A}$ and $\bar{O}_{B}$, where $\bar{O}_{A}<\bar{O}_{B}$, it is necessary first to pass to the disorder parameter set, carry out the operation of subtraction (or division), and map the result back onto the disorganization set, that is,

$$\bar{O}_{A}\rightarrow\bar{O}_{B}=\psi\{\psi^{-1}(\bar{O}_{B})-\psi^{-1}(\bar{O}_{A})+C\} \tag{4.39}$$

or

$$\bar{O}_{A}\rightarrow\bar{O}_{B}=\psi\left\{\frac{\psi^{-1}(\bar{O}_{B})-C_{B}}{\psi^{-1}(\bar{O}_{A})-C_{A}}+C\right\} \tag{4.40}$$

* Here and subsequently, $\cup$ symbolizes union, and $\rightarrow$ symbolizes difference (intersection).

Equations (4.37) and (4.39) are written for the case where the disorder parameter is found by Eq. (4.28), and Eqs. (4.38) and (4.40) for the case where the disorder parameter corresponds to the power disorder, Eq. (4.29).

Expressing changes in the degree of organization in terms of changes in system disorganization under the action of the organizational means, we may write in symbolic form in accordance with Eqs. (4.39) and (4.40)

$$\Delta \bar{O}_i = \bar{O}_{ib} \leftarrow \bar{O}_{ia} \tag{4.41}$$

where $\bar{O}_{ib}$ and $\bar{O}_{ia}$ are the disorganizations before and after the action of organizational means, respectively.

Within the framework of additive interpretation of disorganization, Eq. (4.41) takes the form

$$\Delta \bar{O}_i = \bar{O}_{ib} - \bar{O}_{ia} \tag{4.42}$$

Disorganization may be in terms of position, time, algorithm, structure or statistics. These disorganizations, so widely different in character, may be evaluated by Eqs. (4.31) through (4.35). Basically, the difference is in what is taken as the disorder parameter and how the disorder itself is evaluated [see Eqs. (4.21) through (4.24)].

Let the parameter be the a priori probability, p_j, of occurrence of the jth situation, and the standard of order be $p=1$. Setting $\xi(x)=0$ and substituting the above values in the first line of Eqs. (4.24) gives the statistical power disorder of the jth situation:

$$\bar{Y}_{jp} = 1/p_j - 1 \tag{4.43}$$

Obviously, as $p_j \to 0$, $\bar{Y}_{jp} \to \infty$, while as $p_j \to 1$, $\bar{Y}_{jp} \to 0$. Substituting Eq. (4.43) in Eq. (4.28) and setting $\omega_j = 1$, we get

$$\Pi_{\overline{yjp}} = 1/p_j \tag{4.44}$$

Then we assume ψ to be logarithmic and substitute Eq. (4.44) into (4.33). As a result, we get

$$\bar{O}_p = \sum_{j=1}^{m} p_j \log \frac{1}{p_j} = -\sum_{j=1}^{m} p_j \log p_j$$

which can be recognized as the expression for entropy, Eq. (4.3). Thus, using the general expression for disorganization, we can deduce an expression for statistical disorganization which corresponds to Shannon's entropy.

Consider structural (combinatorial) disorganization. Let the parameter be the number of alternatives, N. Assume that only n of them are optimal, that is, they are standards of order in a sense. Then the disorganization for $\xi(x)=0$ and with an alternative chosen by use of Eq. (4.24) may be written as

$$\bar{Y}_{jk} = N_j/n - 1$$

where $\bar{Y}_{jk} \to \infty$ for $N_j \to \infty$ and $n=$constant, and $\bar{Y}_{jk} \to 0$ for $N_j \to n$. If $n=1$, then

$$\bar{Y}_{jk} = N_j - 1 \tag{4.45}$$

Substituting Eq. (4.45) in Eq. (4.29) at $\omega_j=1$, and then in Eq. (4.33), and also deeming that $m=1$, we obtain

$$\bar{O}_k=\log N$$

which is recognized as the expression deduced by Hartley.

Using Eqs. (4.32) through (4.34), we can deduce several forms of disorganization; that for which the relation ψ is logarithmic will be referred to as *entropy.*

As an example, we shall dwell on goal-associated position entropy

$$H_y=\sum_{i=1}^{k} p_i \log\left[\omega_i(q_i-\xi(x))+1\right] \tag{4.46}$$

It can be deduced by substituting the disorder parameter from Eq. (4.28) in Eq. (4.33), assuming that the relation ψ is logarithmic. Let us interpret this entropy from the stand-point of control applications and analyse its properties.

Let the diversion of the generalized system state vector $\dot{y}$ in the ith realization from the goal vector $\dot{\Psi}$ be designated as the diversion vector $\dot{Q}$. Assume also that all likely diversion realizations, when plotted on the y_1, y_2 coordinates, form a bounded spread space Y which may be decomposed into elementary subspaces $\Delta Y=\Delta^2$, where Δ is the quantizing level, while the expressions log( ) and a( ) symbolize the operations to be performed and have appropriate values in the parentheses (Fig. 4.13). If for each ith elementary subspace we know the a priori probability of finding the vector $\dot{y}$ there and the distance to the goal (this distance q_i is equal to the magnitude of the vector $\dot{Q}$), then for the whole spread space we may write the following diversion ensemble

$$\begin{pmatrix} q_1, & q_2, & \ldots, & q_k \\ p_1, & p_2, & \ldots, & p_k \\ \omega_1, & \omega_2, & \ldots, & \omega_k \end{pmatrix}$$

where $\sum_{i=1}^{k} p_i=1$, and the number of realizations for equal elementary subspaces will be $k=Y/\Delta Y$. With this interpretation, H_y for $\omega_1=\omega_2=\ldots=\omega_k=1$ gives the expected value of the logarithmic diversion from the goal region.

In accordance with Eq. (4.36), for each value of H_y in the entropy space there is an equivalent $H_y(q)$ which represents the expected mean diversion from the goal in the original (spread) space. These two spaces, the entropy space and the original space, are affinely similar (see Fig. 4.13)

The goal-associated position entropy satisfies the following conditions:

(1) $H_y=0$ if all $q\leqslant\xi(x)$ or if $p=0$ for all $q-\xi(x)>0$;

(2) $H_y\to 0$ for the ensemble at $p>0$, if $q-\xi(x)\to 0$, and $H_y\to\infty$ if $q-\xi(x)\to\infty$;

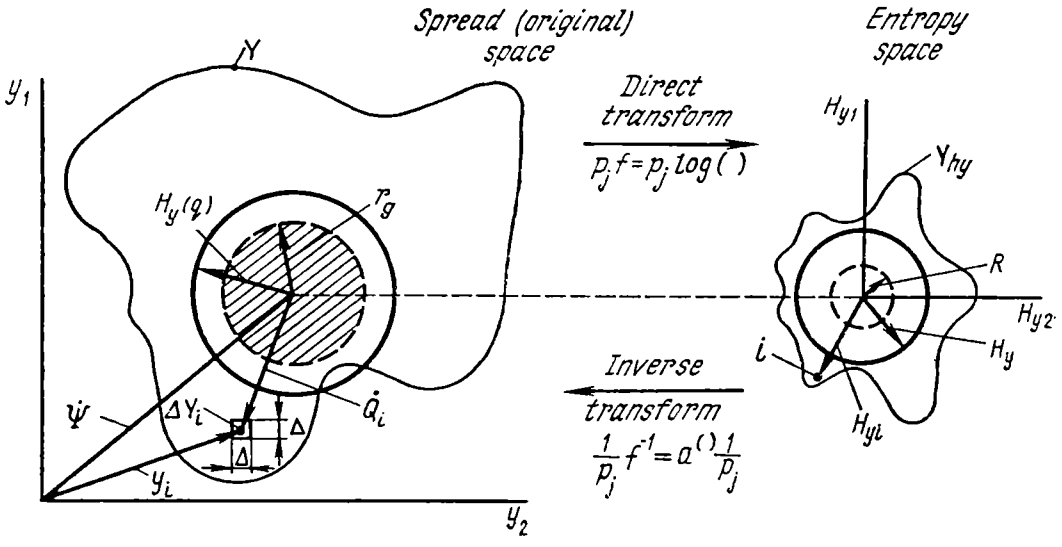


Fig. 4.13

(3) at $p=1/k$,

$$H_y = -\left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{1}{q_i-\xi(x)}\right)+\varepsilon$$

where

$$\varepsilon=\frac{1}{k}\sum_{i=1}^{k}\log\frac{q_i-\xi(x)}{[q_i-\xi(x)]+1}\to 0$$

for $q_i-\xi(x)\gg 1$;

(4) addition of two entropies satisfying the conditions of independence and additivity in the original space gives

$$H_{y(1,2)}=\log\,[H_{y1}(q)+H_{y2}(q)]+\varepsilon$$

where

$$\varepsilon=\log\frac{H_{y1}(q)+H_{y2}(q)}{H_{y1}(q)+H_{y2}(q)+1}\to 0 \quad \text{for} \quad H_y(q)\gg 1$$

Conditions 1 and 2 correspond to intuitive ideas about disorganization in the state of a controlled system with regard to the goal: in the case of ideal control which 'hits' the well-ordered (goal) region in all realizations, $H_y=0$; when the general trend in the ensemble is towards increasing or decreasing the diversion, H_y increases or decreases, respectively.

From condition 3 it follows that in the case of equiprobable diversions, the relation between the entropy space and the spread space with an accuracy up to $\varepsilon\to 0$ at $q\gg 1$ is logarithmic.

Goal-associated entropy is not additive. Goal-associated entropies cannot be added or subtracted directly in the entropy space. To perform these operations, it is necessary to pass to the original space where the condition of additivity is satisfied. After the requisite operations (addition or

subtraction) have been carried out in that space, the next step is to go back to the entropy space.

Condition 4 states a rule by which goal-associated position entropies satisfying the conditions of independence and additivity in the original space can be added with an accuracy of up to $\varepsilon \to 0$ at $H_y(q) \gg 1$.

So far we have dealt with cases involving centred ensembles where the centre of order is located within the spread area so that the major propor-

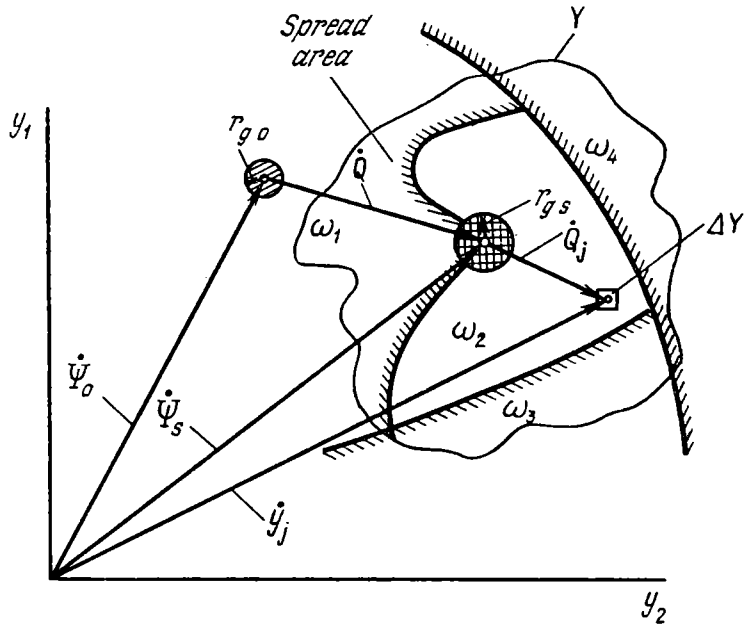


Fig. 4.14

tion (in terms of probability weights) of the ensemble is grouped around it. In the general case, the centre of the ensemble grouping does not coincide with the centre of order. In fact, the latter, as shown in Fig. 4.14, may lie outside of the spread area altogether (which may happen when a system is controlled on the basis of subjective goals different from objective goals). Then position disorganization should be characterized in terms of two components: the mutual position disorganization between the objective goal, $\dot{\Psi}_o$, and the subjective goal, $\dot{\Psi}_s$, and the fractional position entropy defined relative to the subjective goal, that is,

$$H_{(o/s)y} = -\log \frac{1}{g+1} \tag{4.47}$$

$$\Delta H_{sy} = -\sum_{j=1}^{n} p_j \log \frac{1}{q_{(s)j}+1} \tag{4.48}$$

where g is the mutual diversion between the goals, and $q_{(s)j}$ is the spread about the subjective goal.

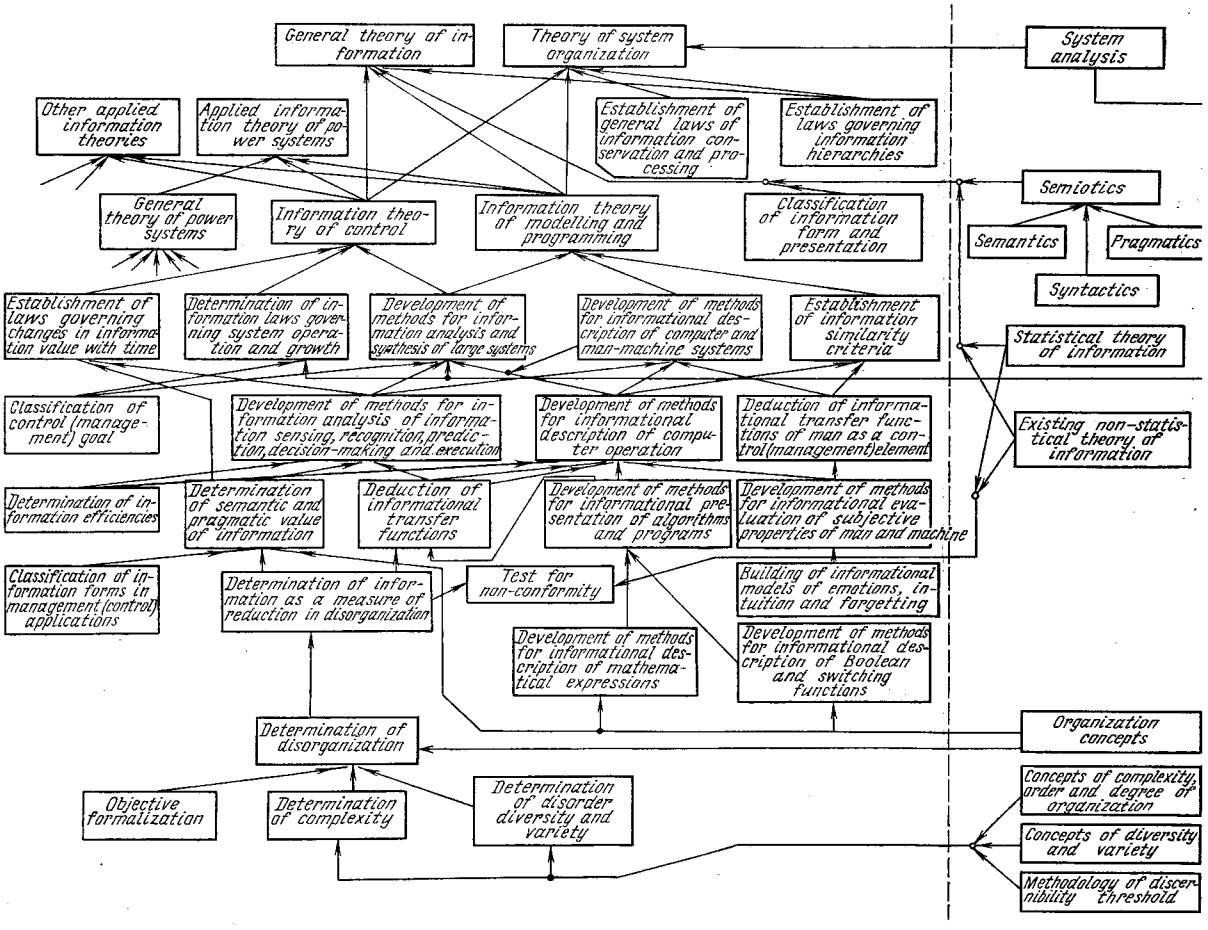

General theory of information
Theory of system organization
System analysis
Other applied information theories
Applied information theory of power systems
Establishment of general laws of information conservation and processing
Establishment of laws governing information hierarchies
Semiotics
General theory of power systems
Information theory of control
Information theory of modelling and programming
Classification of information form and presentation
Semantics
Pragmatics
Syntactics
Establishment of laws governing changes in information value with time
Determination of information laws governing system operation and growth
Development of methods for information analysis and synthesis of large systems
Development of methods for informational description of computer and man-machine systems
Establishment of information similarity criteria
Statistical theory of information
Classification of control (management) goal
Development of methods for information analysis of information sensing, recognition, prediction, decision-making and execution
Development of methods for informational description of computer operation
Deduction of informational transfer functions of man as a control (management) element
Existing non-statistical theory of information
Determination of information efficiencies
Determination of semantic and pragmatic value of information
Deduction of informational transfer functions
Development of methods for informational presentation of algorithms and programs
Development of methods for informational evaluation of subjective properties of man and machine
Classification of information forms in management (control) applications
Determination of information as a measure of reduction in disorganization
Test for non-conformity
Building of informational models of emotions, intuition and forgetting
Development of methods for informational description of mathematical expressions
Development of methods for informational description of Boolean and switching functions
Determination of disorganization
Organization concepts
Objective formalization
Determination of complexity
Determination of disorder diversity and variety
Concepts of complexity, order and degree of organization
Concepts of diversity and variety
Methodology of discernibility threshold

Fig. 4.12

As $g \to 0$, $H_{(o/s)y} \to 0$, and the fractional position entropy satisfies all the four conditions formulated above.

In the subsequent discussion we shall consider a simple case where the centre of order and the centre of the ensemble are the same.

Equation (4.46) has been written on the assumption that q is a continuous quantity occurring as $\Delta Y \to 0$. If the space is quantized into subspaces ΔY (at quantizing levels Δ) which are grouped around the goal vector concentrically, and $\xi(x)=0$ (Fig. 4.15), then the diversions will be uniformly quantified and, as a consequence, the goal-associated entropy will likewise be quantified and it may be defined as

$$H_y = -\sum_{i=1}^{k} p_i \log \frac{1}{N_i+1} \tag{4.49}$$

where N_i is the number of quantizing levels of the discrete diversion $q_i^* = N_i \Delta$.

As will be recalled, Shannon's entropy depends on the size of quantizing levels in such a manner that letting the cardinality of the given set of events

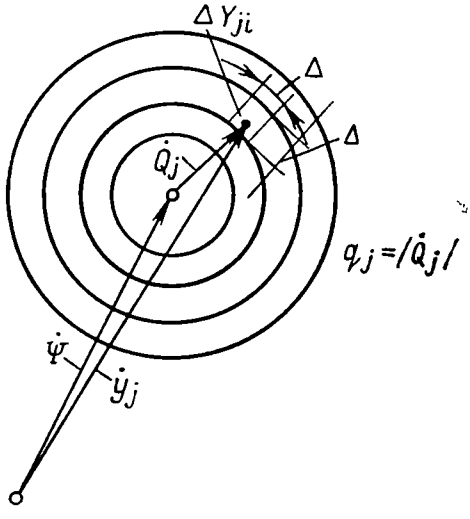


Fig. 4.15

increase to infinity will cause the entropy to tend to infinity, too. This, for example, is true of the continuous interpretation of entropy:

$$H(x) = -\int_{-\infty}^{+\infty} p(x) \log [p(x)\Delta]\, dx$$

where $H(x)$ tends to infinity as $\Delta \to 0$.

In contrast to Shannon's entropy, the goal-associated position entropy defined on a sufficiently small ΔY does not change markedly as ΔY tends to zero. This agrees with the intuitive interpretation which maintains that

goal-associated disorganization should not depend appreciably on the degree of decomposition of the spread space into elementary subspaces.

To give goal-associated position entropy a continuous interpretation, we shall assume (see Fig. 4.15) that the vector $\dot{y}$ has an equal probability p_{jl} of being found in all elementary subspaces ΔY_{jl} removed from the goal by the same distance q_j. Then, for sufficiently small values of Δ, the probability that the vector $\dot{y}$ will be found in the jth annulus can be deduced from the expression known from probability theory:

$$p_{q_j}=(2\pi q_j/\Delta)p_{jl}$$

satisfying

$$\int_0^{q_{\max}} p_q\,dq=1$$

For $\xi(x)=0$, the continuous goal-associated position entropy is defined as

$$H_y=-\int_0^{q_{\max}} p_q \log \frac{1}{q+1}\,dq$$

Since $0\leqslant q\leqslant q_{\max}$, we may change limits by analogy with Shannon's entropy. Then,

$$H_y=-\int_{-\infty}^{+\infty} p_q \log \frac{1}{q+1}\,dq \tag{4.50}$$

In some problems, Shannon's entropy which characterizes the statistical system disorganization for a constant set cardinality, and goal-associated position entropy may be connected by a nonlinear coefficient, $\lambda(H)$. Then a direct relation will immediately be obvious: decreasing H causes H_y to decrease too:

$$H_y=\lambda(H)H \tag{4.51}$$

In the special case where $p_i=1/(q_i+1)$ for all realizations, $H_y=H$. For example, this will be the case if the dispersion of the vector $\dot{y}$ about the goal is constant inside each concentric circle, and the circles are described about the goal so that the probability of finding the vector in the area bounded by the first circle is one-half, in the area between the first and second circles is $1/2^2$, etc., so that the probability of finding the vector between the ith and the $(i+1)$st circles is $p_{i,i+1}=1/2^{i+1}$. The radii will then be connected by the relations: $r_1=1$, $r_2=3r_1$, etc., and $r_i=2^i r_1-1$.

During information transients, changes may occur in the expected value of spread about the goal vector and in the goal region, irrespective of time. In such cases it is convenient, therefore, to deal with the value of entropy H_y at $\xi(x)=0$ (let it be designated as H_{*y}) and a quantity which characterizes the size of the well-ordered (goal) domain in a generalized manner.

If the well-ordered region is arranged symmetrically about the tip of the goal vector, then, by analogy with Eq. (4.46) we may write

$$R = -\log \frac{1}{r_g + 1} \tag{4.52}$$

where r_g is the mean radius of the goal region in the spread space.

In accordance with condition 4, we may write

$$H_y \approx \log_a [H_{*y}(q) - R(q)] = \log_a (a^{H_{*v}} - a^R) \tag{4.53}$$

Since transformation from the spread space into the entropy space utilizes the symbolic operator $f = \log_a(\ ) + 1 \simeq \log_a(\ )$, and the inverse transformation utilizes the operator $f^{-1} = a^{(\ )}$, Eq. (4.53) may be re-written in compact form as

$$H_y = f(f^{-1}H_{*y} - f^{-1}R) \tag{4.54}$$

It is to be noted that in cases involving instantaneous (situational) values (see Fig. 4.13), transformations from the entropy to the spread space and back should be carried out with regard to situational probabilities, that is, use should be made of the operators $p_j f$ and $(1/p_j)f^{-1}$, respectively.

Information as a measure of change in disorganization. In accordance with Eq. (4.41), a general expression for the amount of information may be written as

$$I = \bar{O}_b \leftarrow \bar{O}_a \tag{4.55}$$

Obviously, information will be different according to the form of disorganization considered, its relation to meaning, goal, etc.

In a special case [see Eq. (4.42)], the amount of statistical information is defined as the difference in entropy existing before and after receipt of a message, that is, by Eq. (4.12):

$$I = H_b - H_a$$

Statistical information may be goal-associated, if the disorder parameter is the uncertainty in attaining a specified goal. The goal may be to secure a desired effect with $p = 1$ or to predict objectively occurring probabilistic events with certainty. In the former case, the disorder parameter will be $\Pi_{yj} = 1/p_{aj}$; in the latter case, $\Pi_{yj} = p_j/q_{aj}$, where p_{aj} is the a priori probability of goal attainment, q_{aj} is the a priori forecast probability of a jth event, and p_j is the actual probability of the jth event.

By Eq. (4.33), we get two forms of goal-associated statistical entropy:

$$H_{s.g} = \sum_{j=1}^{n} p_j \log \frac{1}{p_{aj}} \tag{4.56}$$

$$H_{s.g} = \sum_{j=1}^{n} p_j \log \frac{p_j}{q_{aj}} \tag{4.57}$$

In the general case, $p_j \neq p_{aj}$

If the value of information be determined from the difference in entropy before and after receipt of a message, $H_{s.g(b)}$ and $H_{s.g(a)}$, that is, as

$$I_{s.g}=H_{s.g(b)}-H_{s.g(a)}$$

then an expression for the value of information* can be deduced from Eq. (4.56) for $p_1=p_2=\ldots=p_n$, and an expression for the utility of information** by Eq. (4.57).

Speaking about the value of information in control applications, it should be stressed that as any other form of information related to a specific system it carries meaning. Accordingly, it should be examined from two angles. Firstly, it is important to establish how this information reflects the factors that caused its generation (here, information is the effect, and its causes are some processes within the system). Secondly, it is necessary to bring out what changes may be produced in the system when the actions based on this information are applied by controlling means (here information acts as a cause, and the changes in the system as effects). These two aspects of information, as viewed from the stand-point of its utility in determining the state or behaviour of a system and in achieving a particular objective in system operation, will be called the semantic and pragmatic contents of information (leaving out the question of whether this terminology is rigorous enough).

Consider the processing of input information from the most general (language) positions. The processing of information from one language into another may be active or passive. The former occurs when words are translated from language L_{i-1} into language L_i without any change in their meaning. The latter takes place when the meaning of a word translated from language L_{i-1} into language L_i is changed in a purposeful manner (by a certain rule). This purposeful change in meaning takes place, for example, in writing coded messages, when the words in the coded language, L_i, are assigned meanings different from those they have in language L_{i-1} (at least, to a person possessing no key to the code used).

Many control processes may be treated as active translation from a language of images or patterns (input situation) into a language of decision-making.

Conversion of information from one language into another may be direct ($L_{i-1}\rightarrow L_i$) and reverse ($L_i\rightarrow L_{i-1}$). In the case of an ideal translator, passive conversion should leave the amount of utility-weighted information carried by a jth word unchanged in both languages, L_{i-1} and L_i. At the same time, the syntactic amount of information***, determined from the statistical properties of the word or from a combinatorial estimate of the letter sequence forming the jth word, may be different in the two languages. Suppose that the amount of syntactic information carried by a jth word in languages

* For further detail, refer to [1].
** For further detail, refer to [8].
*** See [1, 8].

L_{i-1} and L_i respectively is $I_{i-1,j}$ and $I_{i,j}$, while the amount of information weighted in terms of semantic value is $'I^s_{i-1,j}$ and $'I^s_{i,j}$, and in terms of pragmatic value, $'I^p_{i-1,j}$ and $'I^p_{i,j}$. In the case of passive conversion, the following condition must be satisfied

$$'I^s_{i-1,j}='I^s_{i,j}$$

and

$$'I^p_{i-1,j}='I^p_{i,j}$$

Translation from one language to another may change the value of information, λ*. This change may be expressed as a coefficient of translation, $\alpha_{i/i-1}$, from language L_{i-1} into language L_i. This coefficient defines the extent of change in the value of information per unit of syntactic information. In passive translation, the coefficient of translation will be the same for the semantic and pragmatic value of information:

$$\alpha_{i/i-1,j}=\lambda^s_{i,j}/\lambda^s_{i-1,j}=\lambda^p_{i,j}/\lambda^p_{i-1,j} \tag{4.58}$$

In active translation, the semantic and pragmatic value of information will, in the general case, have different coefficients of translation.

Using the coefficient $\alpha_{i/i-1}$, the following expression may be written for the amount of syntactic information generated in language L_i:

$$I_i=I_{i-1}/\alpha_{i/i-1} \tag{4.59}$$

In reverse translation ($L_i \rightarrow L_{i-1}$), the coefficient of translation is a reciprocal:

$$\alpha_{i-1/i}=1/\alpha_{i/i-1} \tag{4.60}$$

In concrete systems and formulated problems, the semantic and pragmatic values of information are usually interrelated quantities. In control applications, there is a relation between the semantic value of information, that is, in the language of input situations, and the pragmatic value of information at output, that is, in the language of decision-making.

Methods for determining useful (value-weighted) information. As already noted, information can in a wider sense be defined as a measure of the disorganization removed by receipt of a message or, in the general case, by implementation of this message as actions of organizational means. In the general case, information is usefully implementable. In other words, it has not only a potential value available for removal of a disorganization, but it can be put into effect by existing organizational means.

In considering information processing, we have assumed that we have an ideal information processor. In the general case, however, it can both lose some information (generate 'empty' information) and generate some false information.

As already noted, processing of any information by the processor involves the sensing of input information, recognition of situations, prediction, deci-

* The value of information will be defined later.

sion-making, and execution. In each of these stages, information may be expressed in a different language, and the quality of processing will depend on the algorithms used*.

A processing algorithm may yield useful results, L, empty results, L^0, or false results, L^z. Irrespective of the function performed by an algorithm (recognition, decision-making, etc.), its processing properties may in the most general form be expressed as a function of the syntactic amount and value of information about the useful qualities (standard classes) and processing methods

$$L=f(I_{s.c.},\ \lambda^p_{s.c.},\ I_{p.m.},\ \lambda^p_{p.m.}) \tag{4.61}$$

where $I_{s.c.}$ and $I_{p.m.}$ are the syntactic amounts of processing information about standard classes and processing methods, respectively, contained in an algorithm, while $\lambda^p_{s.c.}$ and $\lambda^p_{p.m.}$ are the pragmatic values of this information with regard to the processing goal.

If information processing be thought of as translation from one language, L_{i-1}, into another, L_i, then $L=0$ will signify that the translator has no knowledge of language. Conversely, $L=1$ will signify that the translator has a body of knowledge enabling him to translate words from language L_{i-1} into language L_i ideally. It should be stressed that the translator's knowledge or ignorance are assessed in terms of the words that are used as input information. That is, for $L=1$, it will suffice for the translator to know language to the extent of the words that may arrive at his input. On the other hand, L will be zero, if, no matter how well he may know the language, the translator fails to translate the few words that form the input information set. Thus, the quantity L characterizes the potential capabilities of the translator with regard to a definite input information. Then the pragmatic value of processing information embodied in an algorithm (or, we may say, in a thesaurus of standard classes or a thesaurus of processing methods**) may be defined as

$$\lambda^p_{s.c.}=dL/d[\psi^{-1}(I_{s.c.})] \tag{4.62}$$

$$\lambda^p_{p.m.}=dL/d[\psi^{-1}(I_{p.m.})] \tag{4.63}$$

Apart from the pragmatic properties of an algorithm, that is, its capacity to process information, the information embodied in it can to some extent characterize the receiver's level of training, that is, its semantic properties. Indeed, if we look upon the receiver as a translator, the information stored in his thesaurus will give a measure of his knowledge of some language (or languages). In this, processing information manifests its semantic properties.

---

* Here and elsewhere, an algorithm defines any sequence of steps or a code of rules leading to specific results.

** Here, a *thesaurus* has its conventional meaning of a treasury of knowledge and, especially, the collection of words used by the receiver or translator.

The semantic value of processing information may differ from its pragmatic value. As an example, suppose that a translator is to translate from one language to another a limited number of words such as 'good', 'bad', 'man' and 'machine', which are input information for him. Before long he will be an ideal translator as regards these four words, that is, from a pragmatic point of view. However, his skill in handling these four words tells us little about his knowledge of the languages.

The semantic value of processing information can be assessed in terms of the observer's ignorance about the information processor's (translator's) knowledge, which is removed as the observer learns about the information contained in the translator's thesaurus, be it a thesaurus of standard classes or processing methods. Hence, designating the observer's ignorance as disorganization $\bar{O}_{i\,obs}$, the semantic value may be defined as a derivative

$$\lambda^s_{s.c.}=d[\psi^{-1}(\bar{O}_{i\,obs})]/d[\psi^{-1}(I_{s.c.})] \tag{4.64}$$

$$\lambda^s_{p.m.}=d[\psi^{-1}(\bar{O}_{i\,obs})]/d[\psi^{-1}(I_{p.m.})] \tag{4.65}$$

Apart from processing properties, it is important to take account of the actual implementability of an algorithm by the information processor.

In linguistic interpretation, this means that apart from the potential capabilities of a translator, it is necessary to assess how he utilizes his knowledge in a practical situation. For, the extent to which the translator can apply his knowledge will to a marked extent depend on time constraints, emotional stress, poor retention of words, etc. These factors can be taken care of by introducing the informational coefficient of implementability, η, which will further be referred to as *informational efficiency*.

Assuming steady-state conditions, a general expression for usefully implementable information with regard to the attainment of a pragmatic goal, such as changes in the properties, structure or mix of a controlled system, may be written as

$$'I^p=\Delta\bar{O}_{i(b-a)}=\psi[\psi^{-1}(I_{in})\lambda^p_i L^p_i \eta^p_i] \tag{4.66}$$

A similar expression can be written for usefully implementable information where the goal is of a semantic character, such as the recognition of the properties or behaviour of an information source:

$$'I^s_i=\Delta\bar{O}_{i\,obs\,(b-a)}=\psi[\psi^{-1}(I_{in})\lambda^s_i L^s_i \eta^s_i] \tag{4.67}$$

In Eqs. (4.66) and (4.67):

$\bar{O}_{(b-a)}$, $\bar{O}_{obs\,(b-a)}$ = change in disorganization with regard to goals and subgoals of pragmatic and semantic character, respectively

I_{in} = amount of information (syntactic information is meant) at input to the information processor

λ^p, λ^s = coefficients of pragmatic and semantic value of input information, respectively

L^p, L^s = pragmatic and semantic processing properties of the algorithm used

η^p, η^s = pragmatic and semantic coefficients of implementability of useful information in an information processor with regard to the attainment of a goal

In the above expressions, the value (utility) coefficients are derived on the assumption that $L_i = 1$ and $\eta_i = 1$, from the equations

$$\lambda_i^p = \frac{d[\psi^{-1}(\Delta\bar{O}_{i\,(b-a)})]}{d[\psi^{-1}(I_{in})]}; \qquad \lambda_i^s = \frac{d[\psi^{-1}(\Delta\bar{O}_{i\,obs\,(b-a)})]}{d[\psi^{-1}(I_{in})]} \tag{4.68}$$

Since $0 \leqslant L \leqslant 1$ and $0 \leqslant \eta \leqslant 1$, while λ may take not only zero values (when information is empty), but also negative values (when information is false), both $'I_i$ and $\Delta\bar{O}_i$ may in principle range from $-\infty$ to $+\infty$.

In a special case, setting $\lambda_i = 1$, $\eta_i = 1$ and $\psi(\Pi_y) = \log p$ in Eq. (4.66) or (4.67), we again derive the classical expression for the amount of information, Eq. (4.12).

Consider some special features of informational analysis in control processes.

We assume that the operation of the system considered is defined by a goal, and for each of the above stages in information processing (sensing, recognition, prediction, decision-making, and execution) there are subgoals of their own, with respect to which we shall assess both the value of information and the properties of information processors.

We shall carry out informational analysis on a relative basis — with reference to an isolated control channel, assuming that we are in possession of exhaustive information about the subgoals in all stages of information processing and that the subgoals are the same for all channels. To simplify the matter, we assume that only useful information arrives at the input of the isolated control channel and that its processing does not produce any false information. Also, we assume that all information processes proceed under steady state conditions — there is no accumulation of useful information, no positive feedback between output information and processing properties of algorithms, etc. Under the above assumption, the amount of useful information received can be defined as

$$'I_s = \Delta\bar{O}_s = \psi[\psi^{-1}(I_{in})\lambda_s L_s \eta_s] \tag{4.69}$$

where $\Delta\bar{O}_s$ = reduction in disorganization with regard to the sensing subgoal common to all channels, due to the action of the selected control channel

I_{in}, λ_s = amount of input (syntactic) information arriving at input to the selected channel and its value with regard to the sensing subgoal

L_s = processing properties of the sensing algorithm

η_s = informational efficiency of the reception block

The sensing subgoal may, for example, be to extract only useful part from the total information, and to do so in an amount sufficient to supply

a certain knowledge about the structure, mix and properties of the controlled subsystem.

By analogy with Eq. (4.69), the following expression may be written for the amount of useful information extracted by the recognition block of the selected channel:

$$'I_{rec}=\Delta\bar{O}_{rec}=\psi[\psi^{-1}('I_s)\alpha_{rec/s}L_{rec}\eta_{rec}] \tag{4.70}$$

where $\Delta\bar{O}_{rec}$ = reduction in disorganization with regard to the recognition subgoal due to the action of the selected control channel
$\alpha_{rec/s}$ = coefficient of coupling between the pragmatic value of sensing information and the recognition subgoal
L_{rec} = processing properties of the recognition algorithm
η_{rec} = informational efficiency of the recognition block

Here, too, we have assumed that the recognition subgoal is the same for all control channels.

Reasoning along similar lines, we may write an expression for the amount of useful prediction information which will in the general case be a function of sensing and recognition information. For a large class of problems, prediction information may be represented as the sum of two components associated with sensing information and recognition information:

$$'I_{pred}=\Delta\bar{O}_{pred}=\psi[\psi^{-1}('I_s)\alpha_{pred/s}L_{pred.s}\eta_{pred.s} + \psi^{-1}('I_{rec})\alpha_{pred/rec}L_{pred.rec}\eta_{pred.rec}] \tag{4.71}$$

Decision-making information is a function of prediction information and recognition information. By analogy with Eq. (4.71) we assume that

$$'I_{dm}=\Delta\bar{O}_{dm}=\psi[\psi^{-1}('I_{pred})\alpha_{dm/pred}L_{dm.pred}\eta_{dm.pred} + \psi^{-1}('I_{rec})\alpha_{dm/rec}L_{dm.rec}\eta_{dm.rec}] \tag{4.72}$$

In Eqs. (4.71) and (4.72):

$\alpha_{pred/s}$, $\alpha_{pred/rec}$, $\alpha_{dm/pred}$, $\alpha_{dm/rec}$ = coefficients of coupling between the pragmatic value of sensing, recognition and prediction information and the prediction and decision-making subgoals, respectively

$L_{pred.s}$, $L_{pred.rec}$, $L_{dm.pred}$, $L_{dm.rec}$ = processing properties of the prediction and decision-making algorithms, respectively

$\eta_{pred.s}$, $\eta_{pred.rec}$, $\eta_{dm.pred}$, $\eta_{dm.rec}$ = informational efficiency of the prediction and decision-making blocks, respectively

The decision-making information thus generated is implemented by organizational means as actions minimizing the disorganization of the system with regard to the control goal considered:

$$'I_c=\Delta\bar{O}_c=\psi[\psi^{-1}('I_{dm})\alpha_{c/dm}\eta_c F] \tag{4.73}$$

where $\alpha_{c/dm}$ = coefficient of coupling between the pragmatic value (utility) of decision-making information and the control goal
η_c = informational efficiency of the organizational means with regard to the implementation of $'I_{dm}$
F = indicator of the informational properties of the system*

In a more general case, we may have n nonconflicting control goals and know the coefficients of coupling between the pragmatic value of decision-making information for the selected control channel and each goal ($\alpha_{c1/dm}$, $\alpha_{c2/dm}$, ..., $\alpha_{cn/dm}$) and the informational efficiencies of the organizational means (η_{c1}, η_{c2}, ..., η_{cn}). Then, with regard to each jth goal, we may write

$$\Delta\bar{O}_j = \psi[\psi^{-1}('I_{dm})\alpha_{cj/dm}\eta_{cj}F] \tag{4.74}$$

Apart from pragmatic functions, control information may perform functions of input information; for example, it may characterize the properties of the information processor from semantic positions. A similar remark was made in passing about the semantic value of processing information. However, while the semantic properties of processing information may be examined with regard to one specific algorithm, control information requires that we assess semantically the operation of the entire information processing system, from sensing to decision-making. That is the semantic value of control information must be viewed with regard to the overall performance of the information processor. By assessing control information from semantic positions, we can derive additional input information, I_{in}^{IP} (see Fig. 4.2) about the performance of the processor proper. In this way, we establish information feedback applied to all of the processor.

The semantic approach to the implemented control information establishes its identity with the information existing at the input to the control system, that is, input information. Thus, after it has been handled by the processor, input information is transformed to control information which in turn supplies input information about the processor and, after passage through the control means and the controlled system, input information about the operation of the entire system.

Dynamic properties of information utility, informational transfer functions and informational efficiency. In discussing the parameters of information, we assumed that they were time-invariant or that information was in a steady-state. In fact, this assumption only holds for limited time intervals or for relatively simple problems.

The value of information may degrade with time. This can happen due to at least two causes: firstly, information is devalued at the sink as the user utilizes it; secondly, information becomes obsolete because of delays in transmission or processing. This often happens in practice in both infor-

* Sometimes, such as when the controlled system is a human being, account must be taken of the time and other constraints governing the utilization of useful information by the system.

mation transmission and control. Speaking of the devaluation, or ageing, of information, we shall mean useful (value-weighted) information.

As an example, we shall trace how the information contained in a scientific book is utilized by the human user. Without going into details of the process, we shall assume for simplicity that the extraction of useful information in the time domain is a nonlinear process of the saturation type (curves 'a' in Fig. 4.16), and that it may be defined as

$$'I_{ex}(t) = 'I_0[1 - \exp(-\tau_{ex} t)] \tag{4.75}$$

where $'I_0$, $'I_{ex}(t)$ = total amount of useful information contained in the book and extracted by the reader at time t

τ_{ex} = time constant for extraction of useful information

It is to be noted that in the general case τ_{ex} will depend on the rate at which characters are sensed (number of characters per unit time), that is, on the rate of sensing the syntactic amount of information, I_s; on the arrangement of useful information in the book; the reader's ignorance of the

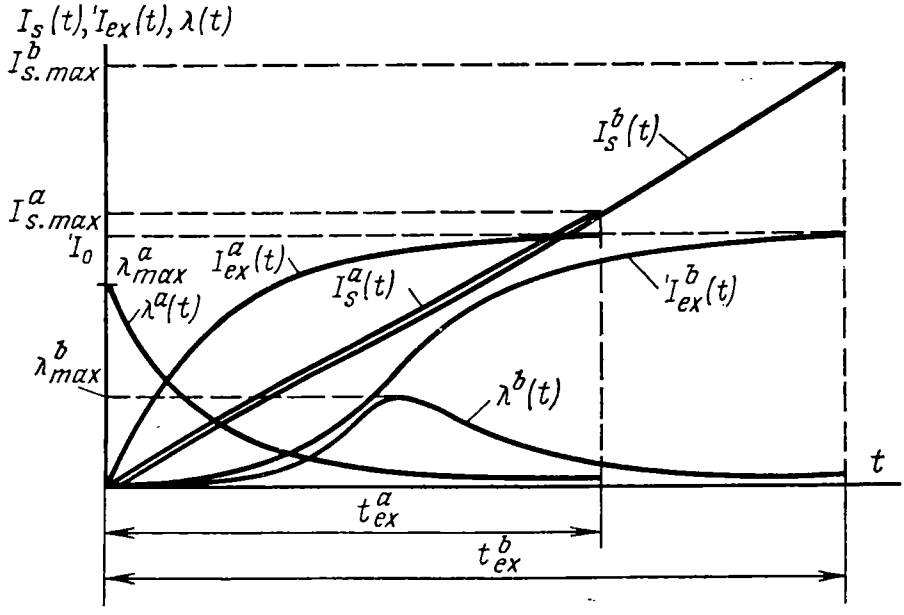


Fig. 4.16

language used in the book (which includes the language of presentation, terms, mathematical formalism, etc.); and on the techniques used to extract useful information.

Assuming that the reader picks up syntactic information at a constant rate (see Fig. 4.16),

$$I_s(t) = kt \tag{4.76}$$

the value of information

$$\lambda = d\{\psi^{-1}['I_{ex}(t)]\}/d\{\psi^{-1}[I_r(t)]\} \tag{4.77}$$

will decrease with time or, it may be said, as the book is read repeatedly.

As another example, consider a control system in which information about the state of the controlled plant is fed over a communication channel. If we allow the delay in the communication channel to increase without bound, we shall finally receive information which ceases to reflect the true state because the plant state will then have changed considerably. For example, information about a fault on a power transmission line will lose its value for fault-clearing devices with increasing delay in the transmission of this information.

This is the ageing of information, because it is assumed that the amount of information (in the syntactic sense) at the input and output of the communication channel remains unchanged. Although in the two examples cited above information was devalued for different reasons, the shape of the curve $\lambda = f(t)$ is the same.

In the general case, the degradation of information is a complex function of not only time, but also the properties of the information source. Yet, for a number of practical problems, this dependence may be stated as an exponential relation of the form:

$$\lambda(t) = \lambda_{\max} \exp(-\tau_{age} t) \tag{4.78}$$

where $\lambda_{\max}$ is the maximum value of information and τ_{age} is the time constant of information ageing.

In some cases, information may gain in value with time. This is true of, say, the utilization of basic research and discoveries. It often happens that at first society either ignores them because of conservatism or cannot put them to practical use because it lacks appropriate technologies. It is not until after society has risen to a higher level of development that these discoveries find use, and the value of information contained in basic research increases. A very illustrative example of this kind is fundamental research in nuclear physics.

Let us go back to the extraction of useful information from a book. Suppose that the reader has not been trained to understand the mathematical formalism used and has therefore to turn to another edition which contains the same amount of useful information (on the subject of interest to the reader) and also an introductory material giving him insight into the formalism used. In such a case, the extraction of useful information from the book may be depicted by curves 'b' in Fig. 4.16. As is seen, the value of information rises at first and falls off at a later time. The run of the curve may be explained by Yu. A. Shreider's theory of the receiver's ignorance.

Another example where information gains in value with time can be taken from control applications using a prediction system. Suppose that prediction is to make up for the delay in the transmission channels or the

delay in the implementation of information by execution means. Obviously, the value of prediction information will increase with increasing anticipation time for the prediction block (provided this increase does not impair the quality of prediction). Prediction information will have a maximum value when the anticipation time is equal to the delay time (which is the condition for information invariance). Thus anticipation is the simplest way to compensate for delay. It may be realized by incorporating derivative action in the excitation control of synchronous generators. With anticipation, the value of information rises over a certain time interval. Information may be said to 'ripen' rather than age. In practical control applications this process may be very complicated, yet approximately it may be assumed that $\lambda(t)$ is an exponential function:

$$\lambda(t)=\lambda_{\max}[1-\exp(-\tau_{ripen}t)] \tag{4.79}$$

where $\lambda_{\max}$ is the maximum value that information can in principle attain in the long run, and τ_{ripen} is the time constant at which information ripens.

At a control centre and in its elements (people, machines) information processing proceeds in time. On the other hand, the processing properties L describe the capabilities of an algorithm with regard to only a particular problem. Even though L may be a time function, it will only show that the processing properties of an algorithm (program) change with time, but it will not describe the dynamics of information handling by the processor. At L=constant the dynamics of information processing depends on the specific qualities of the processor and on the quality with which the algorithm is executed (how the various operations are performed and the program is arranged in storage cells). Therefore, it is important to assess the dynamic properties of the processor elements so as to be able to synthesize it adequately from an informational point of view and to build an optimal hierarchy of computers. Besides, it is to be remembered that man uses information in time, too. Therefore, a proper match between man and machine consists above all in matching their characteristics of information generation and utilization.

Commencement of information processing in a processor or its element, say, a digital computer (after the program and input data have been set in), may be likened to a step application of useful input information, $'\overset{*}{I}_{in}[t]$ to the computer (Fig. 4.17*a*). The response function, $'\overset{*}{I}_{y}[t]$, will depend on the problem being solved, the program used and some other factors (it is assumed that no bound is put on read-out of intermediate results from the computer). The most typical forms of response function appear in Fig. 4.17*b* through *e*. Response *b* is pure delay; in this case information read out of the computer may be defined as

$$'\overset{*}{I}_{yji}[t]=\psi[\psi^{-1}('I_{in}[t_0])L_j\eta_{ji} \tag{4.80}$$

Response *c* is that with a dead zone or dead band; it rises with time after some delay τ. Response *d* is that with saturation or limiting. Response *e*

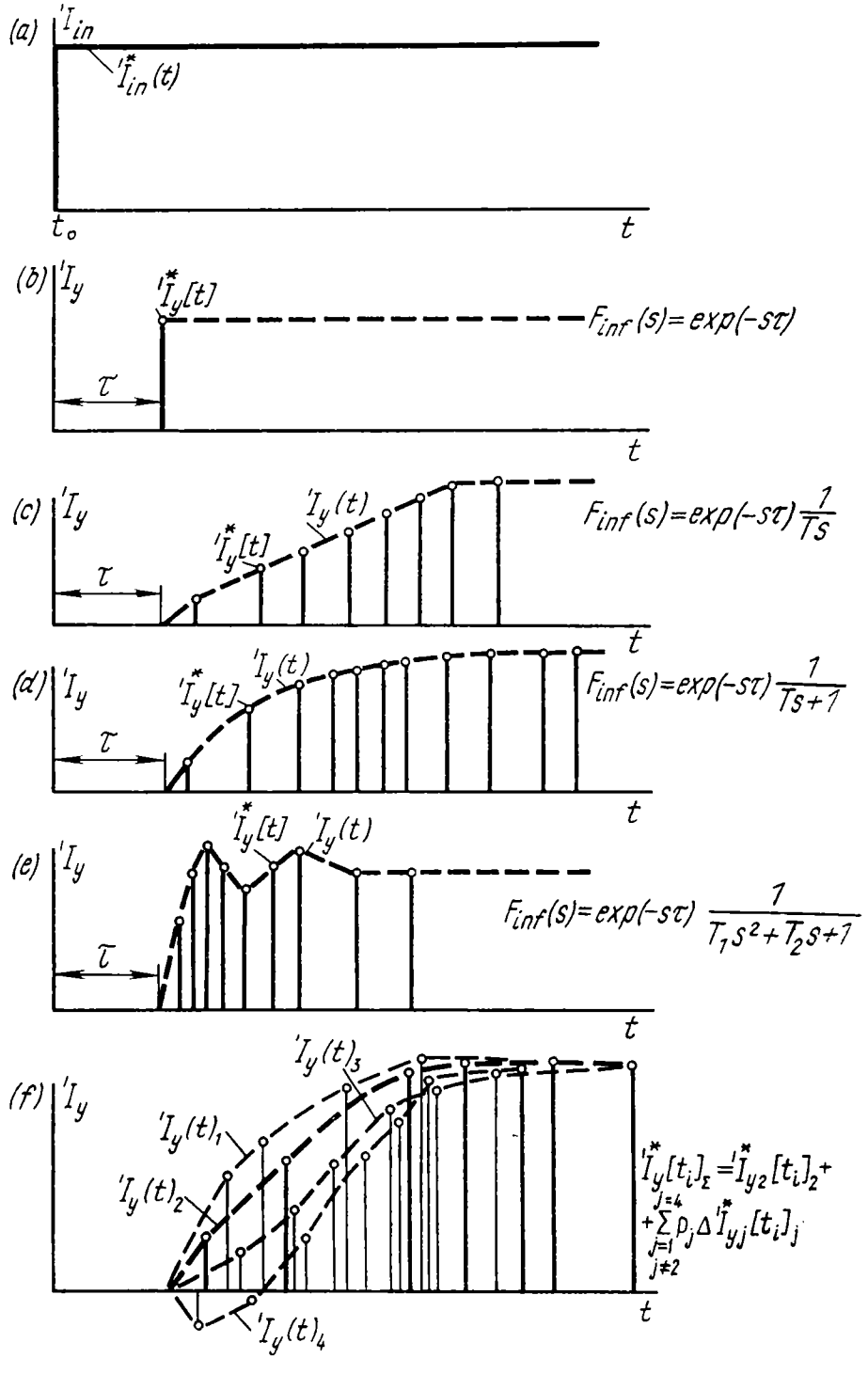

(a)
$'I_{in}$
$'I^*_{in}(t)$
t_o
t
(b)
$'I_y$
$'I^*_y[t]$
$F_{inf}(s)=exp(-s\tau)$
τ
t
(c)
$'I_y$
$'I_y(t)$
$'I^*_y[t]$
$F_{inf}(s)=exp(-s\tau)\frac{1}{Ts}$
τ
t
(d)
$'I_y$
$'I_y(t)$
$'I^*_y[t]$
$F_{inf}(s)=exp(-s\tau)\frac{1}{Ts+1}$
τ
t
(e)
$'I_y$
$'I^*_y[t]$
$'I_y(t)$
$F_{inf}(s)=exp(-s\tau)\frac{1}{T_1s^2+T_2s+1}$
τ
t
(f)
$'I_y$
$'I_y(t)_3$
$'I_y(t)_1$
$'I_y(t)_2$
$'I^*_y[t_i]_\Sigma='I^*_{y2}[t_i]_2+\sum_{\substack{j=1\\j\neq2}}^{j=4}p_j\Delta'I^*_{yj}[t_i]_j$
$'I_y(t)_4$
t

Fig. 4.17

is an oscillatory case. All problems solved by digital computers at present can be reduced to combinations of the first three responses. The fourth type may occur under heuristic programs, such as when, in order to minimize the risk, the computer may discard some alternatives unpredictable as to outcomes and likely to lead to the loss of some useful information.

The dynamic informational properties of a digital computer can be expressed as the ratio of the output function $'I_y[t]$ to the input function $'\overset{*}{I}[t]$, as is done in the theory of automatic control. Since $'\overset{*}{I}_y[t]$ is a discrete function, its rigorous description should be a system of lattice functions. To facilitate the use of the Laplace transform and frequency-response methods, however, it is more convenient to represent the functions in the continuous form.

There exist several techniques by which discrete functions can be converted to continuous. For example, in analysis of the generalized dynamic properties of the digital computer it may be agreed to represent any discrete functions $'\overset{*}{I}_y[t]$ as a continuous envelope $'I_y(t)$. Then, going back to Fig. 4.17*b* through *e*, we may state that, apart from pure delay, the response curves $'I_y(t)$, in the order given, are typical of proportional, integrating, simple-lag and quadratic-lag elements. Assuming that $'I_{in}(t)$ is a step input and using the Laplace transforms for each of the responses, $'I_y(t)$, the informational dynamic properties of the digital computer can be described by appropriate informational transfer functions. In the general case, each will contain a constant term, L, and a variable term, $F_{inf}(s)$:

$$W_{inf}(s)=LF_{inf}(s) \tag{4.81}$$

For the responses depicted in Fig. 4.17*b* through *e*, the transfer functions will respectively take the following forms:

$$F_{inf}(s)=\exp(-s\tau)$$

$$F_{inf}(s)=\exp(-s\tau)\frac{1}{Ts}$$

$$F_{inf}(s)=\exp(-s\tau)\frac{1}{Ts+1}$$

$$F_{inf}(s)=\exp(-s\tau)\frac{1}{T_1s^2+T_2s+1}$$

The transfer functions $F_{inf}(s)$ are valid in the range $0\leqslant'I_y\leqslant'I_{in}L$.

Once the transfer functions are known, the analysis and synthesis of the digital computer as a dynamic system can proceed on the basis of frequency response and other methods employed in the theory of automatic control.

Thus, the digital computer as a dynamic system can be characterized by a transfer function appropriate to the class of problems at hand. However, in contrast to their counterparts used in the theory of automatic control, these functions are informational.

In some cases, such as heuristic, a set of response curves rather than a single $'I_y(t)$ curve results. Before we can determine the transfer function in the circumstances, it is necessary to deduce a generalized response, $'I_y(t)_\Sigma$. This is done by selecting one of the curves as reference (for example, curve *2* in Fig. 4.17*f*) and by determining the difference between the reference curve and each of the remaining ones, $(n-1)$, over the discrete time intervals of interest. Then all of these differences for each *i*th interval are added together according to the probability of occurrence of each curve, to give a generalized curve described by an equation of the form

$$\overset{*}{'I_y}[t_i]_\Sigma = \overset{*}{'I_y}[t_i]_r + \sum_{\substack{j=1 \\ j\neq\delta}}^{n} p_j \Delta \overset{*}{'I_y}[t_i]_j \tag{4.82}$$

where $\overset{*}{'I_y}[t_i]_r$ = value of the reference curve at time t_i

$\Delta \overset{*}{'I}[t_i]_j$ = difference in $'I_y$ between the *j*th and reference curves at time t

p_j = probability that the computer will follow the curve $\overset{*}{'I_y}[t]_j$

With this approach, $\overset{*}{'I_y}[t]_j$ may in the general case take both positive and negative values. In other words, the digital computer may put out false information in the course of operation, say, during a search.

It has earlier been assumed that information processing algorithms remain unchanged as long as the computer is handling a particular problem, because L = constant and $F_{inf}(s)$ = constant. However, these algorithms may change with time within a problem due to changes in the environments or goals. From this stand-point, algorithms may change in accordance with:

(a) a predetermined adaptation algorithm, in which case we have optimization of the processing algorithm, and the algorithms may be called *adaptive;*

(b) algorithms learned by the computer from its 'instructor', in which case we have *learning* algorithms;

(c) algorithms synthesized by the computer itself from its observations of the environments, goals and system performance, in which case we have *self-learning* algorithms.

In operational control, the problems are usually run on a computer by iteration, so that $'I_y(t)$ is changed cyclically, and the time necessary for L and $F_{inf}(s)$ to change is by several orders of magnitude greater than the cyclic time. Therefore, the informational transfer functions of a digital computer should in the general case be divided into those of the information processor proper [see Eq. (4.81)] and the informational transfer functions describing changes in its algorithm. Then the generalized informational transfer function of a digital computer in symbolic notation may be written as

$$\hat{W}_{inf}(s) = L(s)\{[F_{inf}(s)](s)\} \tag{4.83}$$

Assuming that the programmer (which may be a human being or a machine) has an unvarying 'qualification', the informational transfer function of a digital computer is a generalized indicator of computer performance (speed of operation, storage capacity, form of control, etc.) with regard to a particular problem. A similar approach can be used to derive informational transfer functions for the human operator following a prescribed algorithm.

The informational transfer functions of an information processor can be found experimentally from observations of the information generated at its output in response to a step or harmonic change in the amount of input information. In simpler cases, this can be done analytically, by analysing the program informationally.

In the informational analysis demonstrated above, we have taken up only algorithms for information processing. In doing so, we have not differentiated between an algorithm and a program. This is a valid simplification, because informationally there is no difference between an algorithm and a program written in ALGOL, COBOL and similar languages. On the other hand, the specific features of a particular digital computer, such as the range of numbers, storage capacity, speed of operation, etc., are reflected above all in a program written in machine language. From an informational point of view it is of interest to look into matters related to the execution of a program by a digital computer, notably the amount of useful information lost because of limitations on the range of numbers and storage capacity and, in some cases (especially when the same problem is run repeatedly on a computer), the reliability of computer elements (in terms of machine checks and faults). In some problems, limitations on the range of numbers and storage capacity may result in the loss of useful information because of round-offs or too few iteration loops. This loss of information in a digital computer can be assessed in terms of the informational efficiency

$$\eta_{ji}=\psi^{-1}('I_{yji})/\psi^{-1}('I_{yj\max}) \tag{4.84}$$

where $'I_{yj\max}$ is the maximum amount of information useful to the user with regard to the jth problem, which can in principle be generated under the algorithm considered, and $'I_{yji}$ is the information useful to the user with regard to the same problem and under the same algorithm, which is generated at the ith output of the computer in the absence of checks and faults.

In assessing the performance of a digital computer over a long period of time from an informational point of view, it is important to take account of the useful information lost due to machine checks and faults. This can likewise be done in terms of informational efficiency:

$$\eta_{jif}=\sum_{\beta=1}^{n}\psi^{-1}('I_{yjif})_{\beta}\Big/\sum_{\beta=1}^{n}\psi^{-1}('I_{yji})_{\beta} \tag{4.85}$$

where $'I_{yjif}$ is the amount of useful information generated at the output of

a digital computer over n solutions of the problem with account of checks and faults.

In turn, the informational efficiency η_{jif} may be distributed among the three basic units of a digital computer: input, output and processor, as follows:

$$\eta_{jif} = \eta_{jif.\,in}\eta_{jif.\,out}\eta_{jif.\,proc} \tag{4.86}$$

In simple cases, n_{jif} can be evaluated from the failure probabilities of these units.

Relationships between disorganizations in a goal hierarchy. Using the interpretations of goal-related disorganization discussed above, we can tackle one of the most important matters in the informational analysis of large systems, that of the relationships between the goals in a hierarchy. Once these relationships are established, failure (complete or partial) to attain the goals at the lower levels could provide a clue to an estimate of failure in attaining the goals on the higher levels.

Consider the manner in which goals can be defined. If it can be formalized, a goal will in the general case be characterized by the following quantities: a goal vector in some m-ary space of relevant variables, a well-ordered domain with respect to this vector (the feasible region), the safety limit or diversion boundary, the weights assigned to a particular goal to differentiate it from other goals. It may be added that the goal-defining factors may change with time.

It may so happen that a goal thus definable will be attained only partly. In such a case, the system vector will lie outside the well-ordered (feasible) region without crossing the safety limit. The goal attainment indicators may then take continuous or discrete values out of the entire set of values from the boundary of the feasible region to the safety limit.

In the simplest case, a goal may be defined by one-sided or two-sided constraints. In such a case, the boundary of the complete disorder region and the safety limit will be the same, and the degree of goal attainment will take any one of two values, according as the constraints are satisfied or not.

The performance of a system with regard to a specified goal may be assessed in terms of the goal-related disorganization Eqs. (4.47) and (4.48). Then, in the general case, the degree of disorganization with regard to an ith goal can be expressed by a quantity having two components: the amount of mutual disorganization between the centre of the well-ordered region and the centre of the spread ensemble, and relative disorganization, with the centre of the ensemble chosen as reference.

Suppose that the goal vector passes through the centre of the spread ensemble and that the disorganization evaluated with respect to the goal vector is normalized, that is, it can take values from 0 to 1. Then at $\bar{O}=0$ the system will be well-ordered with regard to the specified goal or, which is the same, the goal has been attained completely. At $\bar{O}=1$, the system is completely disordered with regard to that goal; that is, the goal has not been attained.

In goal-related analysis of system performance, the concept of disorganization should be extended to include additional definitions. There may be *characteristic* disorganization and *weighted* disorganization. Characteristic disorganization is found from the assumption that the disorder relevance is decided solely by the amount and probability of disorder. This happens when the observer, in controlling $\overline{Y}_j$ and p_j, does not know how the disorder in the variable concerned affects the performance factors at higher control levels. In such cases, Eq. (4.27) takes the form

$$\bar{O}=\sum_{j=1}^{n} p_j \overline{Y}_j \tag{4.87}$$

In contrast, weighted disorganization [see Eq. (4.27)] comes into the picture when the observer knows how $\overline{Y}_j$ can affect some system performance indicator at a higher control level and, by manipulating ψ and ω, weights $\overline{Y}_j$ in terms of that indicator. The function ψ may be specified axiomatically, but in all cases it must reflect the change in the $\overline{Y}^h$ of the higher-level goal caused by a change in the related disorder of the lower-level goal. In the general case, the weighting procedure should take into account the time factor because $\overline{Y}_j$ may take some time in affecting $\overline{Y}^h$. Also, it may be necessary to reckon with the accumulation effect, etc. This need can be satisfied by introducing an appropriate time function, W_t. Then Eq. (4.27) may be written as

$$\bar{O}=\sum_{j=1}^{n} p_j W_t \psi[\overline{Y}(t_j)] \tag{4.88}$$

where $\overline{Y}(t_j)$ is the disorder in the jth situation with a duration t_j.

In the general case, the dth goal at the $(m+1)$st level may be related to its (say, n) subgoals in the mth level in a very complicated manner. For a large class of systems, however, observed over a time interval T within which the weighting is carried out on the assumption that $W_t=1$, this relationship may be defined as

$$\begin{aligned}\bar{O}_{cd}^{(m+1)}=1-Z_d(1-\gamma_1\bar{O}_{w1}^{(m)})(1-\gamma_2\bar{O}_{w2}^{(m)})\ \ldots\ (1-\gamma_k\bar{O}_{wk}^{(m)})\\ \times(1-\gamma_{k+1}\bar{O}_{w(k+1)}^{(m)}\varrho_{k+1}-\gamma_{k+2}\bar{O}_{w(k+2)}^{(m)}\varrho_{k+2}-\ \ldots\ -\gamma_n\bar{O}_{wn}^{(m)}\varrho_n)\end{aligned} \tag{4.89}$$

where $\bar{O}_{cd}^{(m+1)}$ = characteristic disorganization with regard to the dth goal on the $(m+1)$st level of the goal tree

γ_i = factor of proportionality representing the effect that the dth goal and other subgoals have on the ith subgoal

Z_d = controllability factor of the dth goal

$\bar{O}_{wi}^{(m)}$ = disorganization with regard to the ith subgoal weighted in terms of the indicator of the dth goal

ϱ_i = coefficient of relative effect of $\bar{O}_{wi}^{(m)}$ on $\bar{O}_{cd}^{(m+1)}$

The controllability factor Z_d depends on the structure of relationships between the dth goal and its subgoals and the efficiency of block d respons-

ible for the attainment of the dth goal. In an elementary case, the controllability factor is

$$Z_d = a_1 + a_2 z_{y\,bd} \tag{4.90}$$

where Z_d may take values from 0 to 1. At $Z_d = 1$, the dth goal will be fully attained provided that all of its subgoals have been attained. At $Z_d = 0$, the dth goal cannot be attained even in part for reasons internal to the block concerned (say, a malfunction in the final control elements), although the subgoals have all been attained completely. The quantity $z_{y\,bd}$ defining the efficiency of control block d is in turn a function of three variables

$$z_{y\,bd} = L_{eq.\,i}\eta_{eq.\,i}\eta_{y\,i} \tag{4.91}$$

where $L_{eq.\,i}$ = indicator of the equivalent processing properties of the algorithms used in the ith control block

$\eta_{eq.\,i}$ = equivalent indicator of algorithm implementation by machines and personnel of the ith control block

$\eta_{y\,i}$ = indicator of decision implementation by the final control elements

When the control block shows optimal performance, L_{eq}, η_{eq}, η_y and $z_{y\,b}$ are equal to unity.

The ratio a_2/a_1 is a measure of the responsibility that the control block bears for the attainment of the dth goal. At $a_2/a_1 = 0$, the dth goal is a simple effect of the related subgoals — it is attained each time the related subgoals are attained, with no contribution from the control block.

The attainment of a subgoal may affect that of the related goal in different ways. All subgoals of a goal may be classed into two major groups. Let the subgoals in the first group be called defining [in Eq. (4.90) they are numbered from 1 to k]. If at least one, say the ith, subgoal is not met, that is, if $\bar{O}_{wi}^{(m)} = 1$, then the dth goal cannot be met either. However, the attainment of all defining subgoals cannot guarantee that the dth goal will be fully met. If the dth goal is to be fully attained, it is important to meet all subgoals in the second group, which may arbitrarily be called supplementary [in Eq. (4.90) they are numbered from $(k+1)$ to n]. Failure to meet one or several subgoals in this group will only raise the disorganization of system performance with regard to the dth goal. It may be added in parentheses that if the characteristic disorganization is unity, for the defining subgoals this means that $\bar{O}_{wi}^{(m)} = 1$, too. On the other hand, if $\bar{O}_{ci}^{(m)} = 1$, for the supplementary subgoals this means that $\bar{O}_{wi}^{(m)} \leqslant 1$.

The factor ϱ_i has been introduced in Eq. (4.87) in order to take care of the fact that subgoals are non-additive in their effect on the attainment of the dth goal, that is, the disorganization $\bar{O}_{cd}^{(m+1)}$ is decided not only by the disorganizations of the subgoals, but also by their interaction.

The factor γ_i reflects in a simplified form the fact that the extent to which the attainment of the ith subgoal may affect the attainment of the dth goal varies according to how fully the ith and other subgoals have been attained. In this case, a positive or negative feedback exists between

the degree of attainment of the dth goal and that of the ith subgoal. This factor also reflects the stimulating or suppressing effect of other subgoals on the ith subgoal. Hence, this feedback and cross-coupling factor may be defined as

$$\gamma_i = 1 + \gamma_{id} + \sum_{\substack{l=1 \\ l \neq 1}}^{n} \gamma_{il} \tag{4.92}$$

where γ_{id} is the factor defining the feedback effect that the dth goal has on the ith subgoal, and γ_{il} is the factor defining the cross-effect between the lth and ith subgoals.

When feedback is positive, $\gamma_{id} > 0$. Accordingly, when the lth subgoal has a stimulating effect on the ith subgoal $\gamma_{il} > 0$.

For defining subgoals

$$0 \leqslant \gamma_i \leqslant 1/\bar{O}_{w\,i} \tag{4.93}$$

and for supplementary subgoals

$$0 \leqslant \gamma_i \leqslant \left(1 - \sum_{\substack{r=k+1 \\ r \neq i}}^{n} \gamma_r \bar{O}_{w\,r}^{(m)}\right) \Big/ \bar{O}_{w\,i}^{(m)} \tag{4.94}$$

From Eq. (4.89) it is an easy matter to deduce expressions for other cases of the effect that the attainment of subgoals may have on that of the related goal.

If the ith subgoal is weighted in terms of the dth goal by means of a linear function of the form $\bar{Y}_d = k\bar{Y}_i$ and if $\gamma_i = 1$, then Eq. (4.89) may be re-written as

$$\bar{O}_{cd}^{(m+1)} = 1 - Z_d(1 - \bar{O}_{c1}^{(m)})(1 - \bar{O}_{c2}^{(m)}) \ldots (1 - \bar{O}_{ck}^{(m)}) \times (1 - s_{k+1}\bar{O}_{ck+1}^{(m)} - s_{k+2}\bar{O}_{ck+2}^{(m)} - \ldots - s_n \bar{O}_{cn}^{(m)} \tag{4.95}$$

where $s_i = k_i \varrho_i$ is the relevance factor of the ith subgoal.

If the block responsible for the attainment of the dth goal shows optimal performance ($Z_d = 1$), and the goal has only supplementary or only defining subgoals, Eq. (4.95) may respectively be re-written as

$$\bar{O}_{cd}^{(m+1)} = 1 - \sum_{i=k+1}^{n} s_i \bar{O}_{ci}^{(m)} \tag{4.96}$$

$$\bar{O}_{cd}^{(m+1)} = 1 - \prod_{i=1}^{k} (1 - \bar{O}_{ci}^{(m)}) \tag{4.97}$$

Mathematically, existing techniques for evaluating the effect that the attainment of subgoals has on that of the related goal can be expressed by Eq. (4.96). For example, in the PATTERN (Planning Assistance Through Technical Evaluation of Relevance Number) scheme used in the United States, the relative significance of subgoals is expressed in relevance numbers and it is assumed that subgoals may take values 0 or 1. In operational

control this approach can hardly be used because account must also be taken of a number of other factors, notably the fact that subgoals may be met in part.

In cases where the relevance coefficients cannot be found analytically, resort may be made to experimentation. Suppose that the characteristic disorganizations of the dth goal and of all of its subgoals are known in advance from an analysis of system performance. The appropriate partial derivatives can be deduced from Eq. (4.95):

$$\partial \bar{O}_{cd}^{(m+1)}/d\bar{O}_{ci}^{(m)} = Z_d(1-\bar{O}_{c1}^{(m)})(1-\bar{O}_{c2}^{(m)}) \ldots (1-\bar{O}_{ck}^{(m)})s_i \qquad (4.98)$$

If we assume that disorganizations with regard to the defining subgoals are zero and that $Z_d = 1$, then

$$\partial \bar{O}_{cd}^{(m+1)}/\partial \bar{O}_{ci}^{(m)} = s_i \qquad (4.99)$$

Hence, on replacing Eq. (4.99) by an approximate one, we get

$$s_i = \Delta \bar{O}_{cd}^{(m+1)}/\Delta \bar{O}_{ci}^{(m)} \qquad (4.100)$$

The coefficients s_i give a measure of the sensitivity of $\bar{O}_{cd}^{(m+1)}$ towards changes in $\bar{O}_{ci}^{(m)}$.

Practically, $\bar{O}_{ci}^{(m)}$ can be varied, while keeping constant the disorganizations with regard to the remaining subgoals, by modifying the algorithm or adjusting the amount of information made available to the respective control block or centre. As a result, $\bar{O}_{cd}^{(m+1)}$ will change too, and the s_i's may be found from Eq. (4.100). The effect of subgoals on the attainment of the related goal can be assessed in this way only if $\bar{O}_{cd}^{(m+1)}$ and $\bar{O}_{c}^{(m)}$ are measurable. If the goal defies mathematical formalization, the coefficients s_i can only be found by the expert opinion method.

Equations of information steady and transient states. In terms of goal-related disorganization, systems may be classed into informationally self-regulating and nonself-regulating*. Self-regulating systems are those in which opening the control loop causes goal-related disorganization to increase within certain bounds. An unbounded increase in goal-related disorganization is that when the diversion increases beyond the safety limits of the goal region. Informational self-regulation can be relative, if only one of the control channels is disabled while the others continue in operation. Our further discussion will be limited to informationally self-regulating systems.

Interference, disturbances and other factors causing the vector $\dot{y}$ to veer from the system operation goal may be divided into two categories: *internal*, arising from a variety of internal causes which usually give rise to relatively small and slowly varying diversions (the drift of the vector about the goal), and *external*, arising from external disturbances which may be superim-

* The term self-regulation is used in the theory of automatic control to denote systems capable of maintaining their stability without assistance from automatic control systems. In power engineering, this term is not used.

posed on the drift of the vector $\dot{y}$ and lead to large and rapidly varying diversions. Accordingly, the diversion ensemble may be divided into two independent parts additive in their effects, namely an internal diversion subset and an external diversion subset.

The goal-related disorganization of a system arising after the control loop is opened may be described by re-writing Eqs. (4.37) through (4.40) in symbolic form

$$\bar{O}_* \to \bar{O}_{int} \cup \bar{O}_{ext} \tag{4.101}$$

If disorganization is evaluated by Eqs. (4.37) and (4.39), then Eq. (4.101) can be written out in full as

$$\bar{O}_* = \psi[\psi^{-1}(\bar{O}_{int}) + \psi^{-1}(\bar{O}_{ext}) - \psi^{-1}(R) + C] \tag{4.102}$$

where $\bar{O}_{int}$ = goal-related disorganization caused by internal factors
$\bar{O}_{ext}$ = goal-related disorganization caused by external factors
R = characteristic of the well-ordered region (the region of complete attainment of the goal)

If the control loop is closed and we know the amount and value of input information, the processing properties and informational efficiencies of the information processor in all stages of information processing (sensing, recognition, prediction, decision-making) and the implementability of decision-making information by organizational means, we may evaluate the control information $'I_c$ implemented by the organizational means, using Eqs. (4.69) through (4.74). Then the equation describing the information steady state may be written as

$$\bar{O} = \psi[\psi^{-1}(\bar{O}_*) - \psi^{-1}('I_c) + C] \tag{4.103}$$

At $\bar{O} = 0$, information invariance takes place: $\bar{O}_* = 'I_c$, that is, the information processor and the organizational means cause the system to operate in such a manner that the system vector $\dot{y}$ is found within the well-ordered region (complete attainment of the goal). By the same token, there may be information invariance in each stage of information processing. Also, the information invariance of the entire system may be considered with regard to goal-related disorganization, both internal, $\bar{O}_{int}$, and external, $\bar{O}_{ext}$.

Apart from information invariance, the performance of the information processor and organizational means can be assessed in terms of optimality. It often happens that no clear distinction is drawn between invariance and optimality in control problems. In information problems, this distinction is more definite.

Information invariance is a relative indicator of goal attainment, considered either with regard to the individual stages of information processing or (taking the system as a whole) with regard to particular goals or subgoals within the goal tree.

Information optimality refers to the ideal processing, $L = 1$, of the information available at the input to the processor. In the more general case, this is also the ideal implementation of an information processing algorithm

($\eta = 1$). Similarly, we may speak of the optimality of organizational means, meaning that organization means have an infinite capacity for the control information generated by the information processor. Information optimality may be regarded with reference to both the various stages of information processing and the implementation of control information by organizational means. In the latter case, the optimality of the information processor will be related to the selected goals and subgoals of system operation. In all cases, information optimality, in contrast to information invariance, will be independent of the size of the goal region.

As has been noted more than once, large systems are in constant motion. As they grow and expand, their goal tree and the region of each goal change too, and there is a continuous improvement in the performance of both the information processor and organizational means. This is why an informational study of developing systems must be based on the equation describing the transient information state of the system:

$$\bar{O}(t) = \psi\{\psi^{-1}[\bar{O}_{int}(t)] + \psi^{-1}[\bar{O}_{ext}(t)] - \psi^{-1}[R(t)] - \psi^{-1}['I_c(t)] + C\} \qquad (4.104)$$

In this case, too, we may speak of information invariance and optimality. However, these concepts are more difficult to define under conditions of information transients.

BIBLIOGRAPHY TO CHAPTER 4

1. Temnikov F. E., Afonin V. A., Dmitriev V. I. *Theoretical Fundamentals of Information Engineering* (Teoreticheskie osnovy informatsionnoy tekhniki), "Energiya", 1971.
2. Setrov M. I. *Organization of Biological Systems* (Organizatsiya biosistem). "Nauka", 1971.
3. Gorsky Yu. M., Ursul A. D. *Information in the Management of Large Systems (Mathematical Aspects)* [Informatsiya v upravlenii bolshimi sistemami (matematicheskie aspekty)]. Transaction of the Siberian Power Engineering Institute of the USSR Academy of Sciences. "Optimization and Control of Large Power Systems" (Optimizatsiya i upravlenie v bolshikh sistemakh energetiki). Vol. 3. Irkutsk, 1970.
4. Gorsky Yu. M. *Informational Analysis of Control and Management* (Informatsionny analiz processov upravleniya). Transactions of the Siberian Power Engineering Institute, Siberian Division of the USSR Academy of Sciences. "Optimization and Control of Large Power Systems" (Optimizatsiya i upravlenie v bolshikh sistemakh energetiki). Vol. 1, Irkutsk, 1970.
5. Ursul A. D. *The Nature of Information* (Priroda informatsii). Politizdat, 1968.
6. Venikov V. A., Kengerlinsky G. A. *Information in Power Engineering*. Edited by Academician A. I. Berg. "Sovetskoe Radio", 1967.
7. Petrov B. N., Petrov V. V., Ulanov G. M. et al. *Principles of the Informational Theory of Control*. Engineering Cybernetics. Collected volume "Advances of Science and Technology" (Itogi nauki i tekhniki), 1970.
8. Bongard M. M. *Problems of Recognition* (Problemy uznavaniya). "Nauka", 1967.

INDEX

Alternative, economical, 180
Analysis methods, 34

Branch and bound method, 152

Constraints, discrete, 204
  functional, 212
Criteria, similarity, 183
Criterion analysis, 168
  service reliability, 225
Criterion programming, 230
  fundamental theorem of, 240

Discounted cost equation, 189
Dispatch, economic, 39
Dynamic programming, 135

Ellipsoid of equal precision, 206
Equation, discounted-cost, 189
Expansion forecasting, 46
Extrapolation in forecasting, 63

Forecasting by expert opinion method, 63, 71
  by extrapolation, 63, 65
    by modelling, 63, 79
  expansion, 46
    long-term, 39
      for system expansion, 41
    methods of, 62
    short-term, 39
      for system expansion, 42
    technological, 47
Functions, dual, 235

Gomory's method, 162
Grandient method, 132

Hierarchy, sectoral, 25
  task, 25

Information, basic aspects in control and management, 248
    defined, 245
Information flow, structure of, 265
Information theory of control, 280
Information utility, 298
Informational efficiency, 298
Informational transfer function, 298
Integer programming, 162

Lagrangian functions, modified, 243
Linear programming, 162

Matrix, dimension, 198
Modelling in forecasting, 79
Models, estimation, 106, 131
  mathematical, for power systems, 86
  optimization, 106

Optimality criteria, 168
Optimization, power systems, 54

PATTERN technique, 83
Per unit quantities, 177
Planning, prospective, 39
  technological, 47
Power industry, role in social progress, 9
Power systems, information structure of, 264
    optimization of, 54
Power systems, outline of, 16
Processes, informational, in power systems, 263

Programming, criterion, 230
dynamic, 135
fundamental theorem of, 240
integer, 162
linear, 162
mathematical, 105

Range of variable, 187
Relevance, index of, 83

Similatiry, parametric, 201
Similarity criteria, 183
Simplex algorithm, 107
Simplex method, 115
modified, 123
equations of, 125
State, abnormal, 17
fault, 17
normal, 17
post-fault, 17
State variable, 17
Synthesis methods, 34
System control, operational, 39
short-term, 45
System expansion, economic dispatch in, 44
long-term forecasting of, 41
long-term planning of, 44
optimization of, 101, 105
short-term forecasting of, 42
System variable, 17

Test for economic stability, 175
Theory of similarity, 176
Transportation problem, 140

Units, dimensional, 177
Utility theory, 73

TO THE READER

Mir Publishers welcome your comments on the content, translation, and design of the book.

We would also be pleased to receive any suggestions you care to make about our future publications.

Our address is:
USSR, 129820
Moscow, I-110, GSP,
Pervy Rizhsky Pereulok, 2,
Mir Publishers

Printed in the Union of Soviet Socialist Republics